

STARCH ENCAPSULATION

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to provisional patent application serial No.
 5 60/026,855 filed September 30, 1996. Said provisional application is incorporated herein
 by reference to the extent not inconsistent herewith.

BACKGROUND OF THE INVENTION

Polysaccharide Enzymes

Both prokaryotic and eukaryotic cells use polysaccharide enzymes as a storage
 10 reserve. In the prokaryotic cell the primary reserve polysaccharide is glycogen. Although
 glycogen is similar to the starch found in most vascular plants it exhibits different chain
 lengths and degrees of polymerization. In many plants, starch is used as the primary
 reserve polysaccharide. Starch is stored in the various tissues of the starch bearing plant.
 Starch is made of two components in most instances; one is amylose and one is
 15 amylopectin. Amylose is formed as linear glucans and amylopectin is formed as branched
 chains of glucans. Typical starch has a ratio of 25% amylose to 75% amylopectin.
 Variations in the amylose to amylopectin ratio in a plant can effect the properties of the
 starch. Additionally starches from different plants often have different properties. Maize
 starch and potato starch appear to differ due to the presence or absence of phosphate
 20 groups. Certain plants' starch properties differ because of mutations that have been
 introduced into the plant genome. Mutant starches are well known in maize, rice and peas
 and the like.

The changes in starch branching or in the ratios of the starch components result in
 different starch characteristic. One characteristic of starch is the formation of starch
 25 granules which are formed particularly in leaves, roots, tubers and seeds. These granules
 are formed during the starch synthesis process. Certain synthases of starch, particularly

granule-bound starch synthase, soluble starch synthases and branching enzymes are proteins that are "encapsulated" within the starch granule when it is formed.

The use of cDNA clones of animal and bacterial glycogen synthases are described in International patent application publication number GB92/01881. The nucleotide and amino acid sequences of glycogen synthase are known from the literature. For example, the nucleotide sequence for the *E. coli* glgA gene encoding glycogen synthase can be retrieved from the GenBank/EMBL (SWISSPROT) database, accession number J02616 (Kumar et al., 1986, J. Biol. Chem., 261:16256-16259). *E. coli* glycogen biosynthetic enzyme structural genes were also cloned by Okita et al. (1981, J. Biol. Chem., 256(13):6944-6952). The glycogen synthase glgA structural gene was cloned from *Salmonella typhimurium* LT2 by Leung et al. (1987, J. Bacteriol., 169(9):4349-4354). The sequences of glycogen synthase from rabbit skeletal muscle (Zhang et al., 1989, FASEB J., 3:2532-2536) and human muscle (Browner et al., 1989, Proc. Natl. Acad. Sci., 86:1443-1447) are also known.

The use of cDNA clones of plant soluble starch synthases has been reported. The amino acid sequences of pea soluble starch synthase isoforms I and II were published by Dry et al. (1991, Plant Journal, 2:193202). The amino acid sequence of rice soluble starch synthase was described by Baba et al. (1993, Plant Physiology,). This last sequence (rice SSTS) incorrectly cites the N-terminal sequence and hence is misleading. Presumably this is because of some extraction error involving a protease degradation or other inherent instability in the extracted enzyme. The correct N-terminal sequence (starting with AELSR) is present in what they refer to as the transit peptide sequence of the rice SSTS.

The sequence of maize branching enzyme I was investigated by Baba et al., 1991, BBRC, 181:8794. Starch branching enzyme II from maize endosperm was investigated by Fisher and Shrable (1993, Plant Physiol., 102:10451046). The use of cDNA clones of plant, bacterial and animal branching enzymes have been reported. The nucleotide and amino acid sequences for bacterial branching enzymes (BE) are known from the literature. For example, Kiel et al. cloned the branching enzyme gene glgB from *Cyanobacterium synechococcus* PCC7942 (1989, Gene (Amst), 78(1):918) and from *Bacillus*

stearothermophilus (Kiel et al., 1991, Mol. Gen. Genet., 230(12):136-144). The genes *glc3* and *ghal* of *S. cerevisiae* are allelic and encode the glycogen branching enzyme (Rowen et al., 1992, Mol. Cell Biol., 12(1):22-29). Matsumomoto et al. investigated glycogen branching enzyme from *Neurospora crassa* (1990, J. Biochem., 107:118-122).
5 The GenBank/EMBL database also contains sequences for the *E. coli* *glgB* gene encoding branching enzyme.

Starch synthase (EC 2.4.1.11) elongates starch molecules and is thought to act on both amylose and amylopectin. Starch synthase (STS) activity can be found associated both with the granule and in the stroma of the plastid. The capacity for starch association
10 of the bound starch synthase enzyme is well known. Various enzymes involved in starch biosynthesis are now known to have differing propensities for binding as described by Mu-Forster et al. (1996, Plant Phys. 111: 821-829). Granule-bound starch synthase (GBSTS) activity is strongly correlated with the product of the *waxy* gene (Shure et al., 1983, Cell 35: 225-233). The synthesis of amylose in a number of species such as maize, rice and
15 potato has been shown to depend on the expression of this gene (Tsai, 1974, Biochem Gen 11: 83-96; Hovenkamp-Hermelink et al., 1987, Theor. Appl. Gen. 75: 217-221). Visser et al. described the molecular cloning and partial characterization of the gene for granule-bound starch synthase from potato (1989, Plant Sci. 64(2):185-192). Visser et al. have also described the inhibition of the expression of the gene for granule-bound starch
20 synthase in potato by antisense constructs (1991, Mol. Gen. Genet. 225(2):289-296).

The other STS enzymes have become known as soluble starch synthases, following the pioneering work of Frydman and Cardini (Frydman and Cardini, 1964, Biochem. Biophys. Res. Communications 17: 407-411). Recently, the appropriateness of the term "soluble" has become questionable in light of discoveries that these enzymes are
25 associated with the granule as well as being present in the soluble phase (Denyer et al., 1993, Plant J. 4: 191-198; Denyer et al., 1995, Planta 97: 57-62; Mu-Forster et al., 1996, Plant Physiol. 111: 821-829). It is generally believed that the biosynthesis of amylopectin involves the interaction of soluble starch synthases and starch branching enzymes. Different isoforms of soluble starch synthase have been identified and cloned in pea
30 (Denyer and Smith, 1992, Planta 186: 609-617; Dry et al., 1992, Plant Journal, 2: 193-

202), potato (Edwards et al., 1995, Plant Physiol 112: 89-97; Marshall et al., 1996, Plant Cell 8: 1121-1135) and in rice (Baba et al., 1993, Plant Physiol. 103: 565-573), while barley appears to contain multiple isoforms, some of which are associated with starch branching enzyme (Tyynela and Schulman, 1994, Physiol. Plantarum 89: 835-841). A
5 common characteristic of STS clones is the presence of a KXGGLGDV consensus sequence which is believed to be the ADP-Glc binding site of the enzyme (Furukawa et al., 1990, J Biol Chem 265: 2086-2090; Furukawa et al., 1993, J. Biol. Chem. 268: 23837-23842).

In maize, two soluble forms of STS, known as isoforms I and II, have been
10 identified (Macdonald and Preiss, 1983, Plant Physiol. 73: 175-178; Boyer and Preiss, 1978, Carb. Res. 61: 321-334; Pollock and Preiss, 1980, Arch Biochem. Biophys. 204: 578-588; Macdonald and Preiss, 1985 Plant Physiol. 78: 849-852; Dang and Boyer, 1988, Phytochemistry 27: 1255-1259; Mu et al., 1994, Plant J. 6: 151-159), but neither of these has been cloned. STSI activity of maize endosperm was recently correlated with a 76-kDa
15 polypeptide found in both soluble and granule-associated fractions (Mu et al., 1994, Plant J. 6: 151-159). The polypeptide identity of STSII remains unknown. STSI and II exhibit different enzymological characteristics. STSI exhibits primer-independent activity whereas STSII requires glycogen primer to catalyze glucosyl transfer. Soluble starch synthases
20 have been reported to have a high flux control coefficient for starch deposition (Jenner et al., 1993, Aust. J. Plant Physiol. 22: 703-709; Keeling et al., 1993, Planta 191: 342-348) and to have unusual kinetic properties at elevated temperatures (Keeling et al., 1995, Aust. J. Plant Physiol. 21 807-827). The respective isoforms in maize exhibit significant differences in both temperature optima and stability.

Plant starch synthase (and *E. coli* glycogen synthase) sequences include the
25 sequence KTGGL which is known to be the ADPG binding domain. The genes for any such starch synthase protein may be used in constructs according to this invention.

Branching enzyme [α 1,4Dglucan: α 1,4Dglucan 6D(α 1,4Dglucano) transferase (E.C. 2.4.1.18)], sometimes called Q-enzyme, converts amylose to amylopectin. A segment of a α 1,4Dglucan chain is transferred to a primary hydroxyl group in a similar glucan chain.

Bacterial branching enzyme genes and plant sequences have been reported (rice endosperm: Nakamura et al., 1992, *Physiologia Plantarum*, 84:329-335 and Nakamura and Yamanouchi, 1992, *Plant Physiol.*, 99:1265-1266; pea: Smith, 1988, *Planta*, 175:270-279 and Bhattacharyya et al., 1989, *J. Cell Biochem., Suppl.* 13D:331; maize endosperm: Singh and Preiss, 1985, *Plant Physiology*, 79:34-40; VosScherperkeuter et al., 1989, *Plant Physiology*, 90:75-84; potato: Kossmann et al., 1991, *Mol. Gen. Genet.*, 230(12):39-44; cassava: Salehuzzaman and Visser, 1992, *Plant Mol Biol*, 20:809-819).

In the area of polysaccharide enzymes there are reports of vectors for engineering modification in the starch pathway of plants by use of a number of starch synthesis genes in various plant species. That some of these polysaccharide enzymes bind to cellulose or starch or glycogen is well known. One specific patent example of the use of a polysaccharide enzyme shows the use of glycogen biosynthesis enzymes to modify plant starch. In U.S. patent 5,349,123 to Shewmaker a vector containing DNA to form glycogen biosynthetic enzymes within plant cells is taught. Specifically, this patent refers to the changes in potato starch due to the introduction of these enzymes. Other starch synthesis genes and their use have also been reported.

Hybrid (fusion) Peptides

Hybrid proteins (also called "fusion proteins") are polypeptide chains that consist of two or more proteins fused together into a single polypeptide. Often one of the proteins is a ligand which binds to a specific receptor cell. Vectors encoding fusion peptides are primarily used to produce foreign proteins through fermentation of microbes. The fusion proteins produced can then be purified by affinity chromatography. The binding portion of one of the polypeptides is used to attach the hybrid polypeptide to an affinity matrix. For example, fusion proteins can be formed with beta galactosidase which can be bound to a column. This method has been used to form viral antigens.

Another use is to recover one of the polypeptides of the hybrid polypeptide. Chemical and biological methods are known for cleaving the fused peptide. Low pH can be used to cleave the peptides if an acid-labile aspartyl-proline linkage is employed between the peptides and the peptides are not affected by the acid. Hormones have been

cleaved with cyanobromide. Additionally, cleavage by site-specific proteolysis has been reported. Other methods of protein purification such as ion chromatography have been enhanced with the use of polyarginine tails which increase overall basicity of the protein thus enhancing binding to ion exchange columns.

5 A number of patents have outlined improvements in methods of making hybrid peptides or specific hybrid peptides targeted for specific uses. US patent 5,635,599 to Pastan et al. outlines an improvement of hybrid proteins. This patent reports a circularly permuted ligand as part of the hybrid peptide. This ligand possesses specificity and good binding affinity. Another improvement in hybrid proteins is reported in U.S. patent
10 5,648,244 to Kuliopulos. This patent describes a method for producing a hybrid peptide with a carrier peptide. This nucleic acid region, when recognized by a restriction endonuclease, creates a nonpalindromic 3-base overhang. This allows the vector to be cleaved.

 An example of a specifically targeted hybrid protein is reported in U.S. patent
15 5,643,756. This patent reports a vector for expression of glycosylated proteins in cells. This hybrid protein is adapted for use in proper immunoreactivity of HIV gp120. The isolation of gp120 domains which are highly glycosylated is enhanced by this reported vector.

 U.S. patent 5,202,247 and 5,137,819 discuss hybrid proteins having polysaccharide
20 binding domains and methods and compositions for preparation of hybrid proteins which are capable of binding to a polysaccharide matrix. U.S. patent 5,202,247 specifically teaches a hybrid protein linking a cellulase binding region to a peptide of interest. The patent specifies that the hybrid protein can be purified after expression in a bacterial host by affinity chromatography on cellulose.

25 The development of genetic engineering techniques has made it possible to transfer genes from various organisms and plants into other organisms or plants. Although starch has been altered by transformation and mutagenesis in the past there is still a need for further starch modification. To this end vectors that provide for encapsulation of desired

amino acids or peptides within the starch and specifically within the starch granule are desirable. The resultant starch is modified and the tissue from the plant carrying the vector is modified.

SUMMARY OF THE INVENTION

5 This invention provides a hybrid polypeptide comprising a starch-encapsulating region (SER) from a starch-binding enzyme fused to a payload polypeptide which is not endogenous to said starch-encapsulating region, i.e. does not naturally occur linked to the starch-encapsulating region. The hybrid polypeptide is useful to make modified starches comprising the payload polypeptide. Such modified starches may be used to provide grain
10 feeds enriched in certain amino acids. Such modified starches are also useful for providing polypeptides such as hormones and other medicaments, e.g. insulin, in a starch-encapsulated form to resist degradation by stomach acids. The hybrid polypeptides are also useful for producing the payload polypeptides in easily-purified form. For example, such hybrid polypeptides produced by bacterial fermentation, or in grains or animals, may
15 be isolated and purified from the modified starches with which they are associated by art-known techniques.

The term "polypeptide" as used herein means a plurality of identical or different amino acids, and also encompasses proteins.

20 The term "hybrid polypeptide" means a polypeptide composed of peptides or polypeptides from at least two different sources, e.g. a starch-encapsulating region of a starch-binding enzyme, fused to another polypeptide such as a hormone, wherein at least two component parts of the hybrid polypeptide do not occur fused together in nature.

25 The term "payload polypeptide" means a polypeptide not endogenous to the starch-encapsulating region whose expression is desired in association with this region to express a modified starch containing the payload polypeptide.

When the payload polypeptide is to be used to enhance the amino acid content of particular amino acids in the modified starch, it preferably consists of not more than three different types of amino acids selected from the group consisting of: Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, and Val.

5 When the payload polypeptide is to be used to supply a biologically active polypeptide to either the host organism or another organism, the payload polypeptide may be a biologically active polypeptide such as a hormone, e.g., insulin, a growth factor, e.g. somatotropin, an antibody, enzyme, immunoglobulin, or dye, or may be a biologically active fragment thereof as is known to the art. So long as the polypeptide has biological activity, it does not need to be a naturally-occurring polypeptide, but may be mutated, 10 truncated, or otherwise modified. Such biologically active polypeptides may be modified polypeptides, containing only biologically-active portions of biologically-active polypeptides. They may also be amino acid sequences homologous to naturally-occurring biologically-active amino acid sequences (preferably at least about 75% homologous) 15 which retain biological activity.

The starch-encapsulating region of the hybrid polypeptide may be a starch-encapsulating region of any starch-binding enzyme known to the art, e.g. an enzyme selected from the group consisting of soluble starch synthase I, soluble starch synthase II, soluble starch synthase III, granule-bound starch synthase, branching enzyme I, branching 20 enzyme IIa, branching enzyme IIBb and glucoamylase polypeptides.

When the hybrid polypeptide is to be used to produce payload polypeptide in pure or partially purified form, the hybrid polypeptide preferably comprises a cleavage site between the starch-encapsulating region and the payload polypeptide. The method of isolating the purified payload polypeptide then includes the step of contacting the hybrid 25 polypeptide with a cleaving agent specific for that cleavage site.

This invention also provides recombinant nucleic acid (RNA or DNA) molecules encoding the hybrid polypeptides. Such recombinant nucleic acid molecules preferably comprise control sequences adapted for expression of the hybrid polypeptide in the

selected host. The term "control sequences" includes promoters, introns, preferred codon sequences for the particular host organism, and other sequences known to the art to affect expression of DNA or RNA in particular hosts. The nucleic acid sequences encoding the starch-encapsulating region and the payload polypeptide may be naturally-occurring
5 nucleic acid sequences, or biologically-active fragments thereof, or may be biologically-active sequences homologous to such sequences, preferably at least about 75% homologous to such sequences.

Host organisms include bacteria, plants, and animals. Preferred hosts are plants. Both monocotyledonous plants (monocots) and dicotyledonous plants (dicots) are useful
10 hosts for expressing the hybrid polypeptides of this invention.

This invention also provides expression vectors comprising the nucleic acids encoding the hybrid proteins of this invention. These expression vectors are used for transforming the nucleic acids into host organisms and may also comprise sequences aiding in the expression of the nucleic acids in the host organism. The expression vectors
15 may be plasmids, modified viruses, or DNA or RNA molecules, or other vectors useful in transformation systems known to the art.

By the methods of this invention, transformed cells are produced comprising the recombinant nucleic acid molecules capable of expressing the hybrid polypeptides of this invention. These may prokaryotic or eukaryotic cells from one-celled organisms, plants or
20 animals. They may be bacterial cells from which the hybrid polypeptide may be harvested. Or, they may be plant cells which may be regenerated into plants from which the hybrid polypeptide may be harvested, or, such plant cells may be regenerated into fertile plants with seeds containing the nucleic acids encoding the hybrid polypeptide. In a preferred embodiment, such seeds contain modified starch comprising the payload
25 polypeptide.

The term "modified starch" means the naturally-occurring starch has been modified to comprise the payload polypeptide.

A method of targeting digestion of a payload polypeptide to a particular phase of the digestive process, e.g., preventing degradation of a payload polypeptide in the stomach of an animal, is also provided comprising feeding the animal a modified starch of this invention comprising the payload polypeptide, whereby the polypeptide is protected by the starch from degradation in the stomach of the animal. Alternatively, the starch may be one known to be digested in the stomach to release the payload polypeptide there.

Preferred recombinant nucleic acid molecules of this invention comprise DNA encoding starch-encapsulating regions selected from the starch synthesizing gene sequences set forth in the tables hereof.

Preferred plasmids of this invention are adapted for use with specific hosts. Plasmids comprising a promoter, a plastid-targeting sequence, a nucleic acid sequence encoding a starch-encapsulating region, and a terminator sequence, are provided herein. Such plasmids are suitable for insertion of DNA sequences encoding payload polypeptides and starch-encapsulating regions for expression in selected hosts.

Plasmids of this invention can optionally include a spacer or a linker unit proximate the fusion site between nucleic acids encoding the SER and the nucleic acids encoding the payload polypeptide. This invention includes plasmids comprising promoters adapted for a prokaryotic or eukaryotic hosts. Such promoters may also be specifically adapted for expression in monocots or in dicots.

A method of forming peptide-modified starch of this invention includes the steps of: supplying a plasmid having a promoter associated with a nucleic acid sequence encoding a starch-encapsulating region, the nucleic acid sequence encoding the starch-encapsulating region being connected to a nucleic acid region encoding a payload polypeptide, and transforming a host with the plasmid whereby the host expresses peptide-modified starch.

This invention furthermore comprises starch-bearing grains comprising: an embryo, nutritive tissues; and, modified starch granules having encapsulated therein a protein that is

not endogenous to starch granules of said grain which are not modified. Such starch-bearing grains may be grains wherein the embryo is a maize embryo, a rice embryo, or a wheat embryo.

5 All publications referred to herein are incorporated by reference to the extent not inconsistent herewith.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1a shows the plasmid pEXS114 which contains the synthetic GFP (Green Fluorescent Protein) subcloned into pBSK from Stratagene.

FIG. 1b shows the plasmid pEXS115.

10 FIG. 2a. shows the *waxy* gene with restriction sites subcloned into a commercially available plasmid.

FIG. 2b shows the p ET-21A plasmid commercially available from Novagen having the GFP fragment from pEXS115 subcloned therein.

FIG. 3a shows pEXS114 subcloned into pEXSWX, and the GFP-FLWX map.

15 FIG. 3b shows the GFP-Bam HIWX plasmid.

FIG. 4 shows the SGFP fragment of pEXS115 subcloned into pEXSWX, and the GFP-NcoWX map.

FIG. 5 shows a linear depiction of a plasmid that is adapted for use in monocots.

FIG. 6 shows the plasmid pEXS52.

FIG. 7 shows the six introductory plasmids used to form pEXS51 and pEX560. FIG. 7a shows pEXS adh1. FIG. 7b shows pEXS adh1-nos3'. FIG. 7c shows pEXS33. FIG. 7d shows pEXS10zp. FIG. 7e shows pEXS10zp-adh1. FIG. 7f shows pEXS10zp-adh1-nos3'.

FIGS. 8a and 8b show the plasmids pEXS50 and pEXS51, respectively, containing the MS-SIII gene which is a starch-soluble synthase gene.

FIG. 9a shows the plasmid pEXS60 which excludes the intron shown in pEXS50, and FIG. 9b shows the plasmid pEXS61 which excludes the intron shown in pEXS60.

DETAILED DESCRIPTION

The present invention provides, broadly, a hybrid polypeptide, a method for making a hybrid polypeptide, and nucleic acids encoding the hybrid polypeptide. A hybrid polypeptide consists of two or more subparts fused together into a single peptide chain. The subparts can be amino acids or peptides or polypeptides. One of the subparts is a starch-encapsulating region. Hybrid polypeptides may thus be targeted into starch granules produced by organisms expressing the hybrid polypeptides.

A method of making the hybrid polypeptides within cells involves the preparation of a DNA construct comprising at least a fragment of DNA encoding a sequence which functions to bind the expression product of attached DNA into a granule of starch, ligated to a DNA sequence encoding the polypeptide of interest (the payload polypeptide). This construct is expressed within a eukaryotic or prokaryotic cell. The hybrid polypeptide can be used to produce purified protein or to immobilize a protein of interest within the protection of a starch granule, or to produce grain that contains foreign amino acids or peptides.

The hybrid polypeptide according to the present invention has three regions.

Payload Peptide (X)	Central Site (CS)*	Starch-encapsulating region (SER)
------------------------	-----------------------	--------------------------------------

X is any amino acid or peptide of interest.

* optional component.

The gene for X can be placed in the 5' or 3' position within the DNA construct described below.

5 CS is a central site which may be a leaving site, a cleavage site, or a spacer, as is known to the art. A cleavage site is recognized by a cleaving enzyme. A cleaving enzyme is an enzyme that cleaves peptides at a particular site. Examples of chemicals and enzymes that have been employed to cleave polypeptides include thrombin, trypsin, cyanobromide, formic acid, hydroxyl amine, collagenase, and alacubtilisin. A spacer is a
10 peptide that joins the peptides comprising the hybrid polypeptide. Usually it does not have any specific activity other than to join the peptides or to preserve some minimum distance or to influence the folding, charge or water acceptance of the protein. Spacers may be any peptide sequences not interfering with the biological activity of the hybrid polypeptide.

The starch-encapsulating region (SER) is the region of the subject polypeptide that
15 has a binding affinity for starch. Usually the SER is selected from the group consisting of peptides comprising starch-binding regions of starch synthases and branching enzymes of plants, but can include starch binding domains from other sources such as glucoamylase and the like. In the preferred embodiments of the invention, the SER includes peptide products of genes that naturally occur in the starch synthesis pathway. This subset of
20 preferred SERs is defined as starch-forming encapsulating regions (SFER). A further subset of SERs preferred herein is the specific starch-encapsulating regions (SSER) from the specific enzymes starch synthase (STS), granule-bound starch synthase (GBSTS) and branching enzymes (BE) of starch-bearing plants. The most preferred gene product from this set is the GBSTS. Additionally, starch synthase I and branching enzyme II are useful
25 gene products. Preferably, the SER (and all the subsets discussed above) are truncated versions of the full length starch synthesizing enzyme gene such that the truncated portion includes the starch-encapsulating region.

The DNA construct for expressing the hybrid polypeptide within the host, broadly is as follows:

Promoter	Intron*	Transit Peptide Coding Region*	X	SER	Terminator
-----------------	----------------	---	----------	------------	-------------------

* optional component. Other optional components can also be used.

5 As is known to the art, a promoter is a region of DNA controlling transcription. Different types of promoters are selected for different hosts. Lac and T7 promoters work well in prokaryotes, the 35S CaMV promoter works well in dicots, and the polyubiquitin promoter works well in many monocots. Any number of different promoters are known to the art and can be used within the scope of this invention.

10 Also as is known to the art, an intron is a nucleotide sequence in a gene that does not code for the gene product. One example of an intron that often increases expression in monocots is the Adh1 intron. This component of the construct is optional.

15 The transit peptide coding region is a nucleotide sequence that encodes for the translocation of the protein into organelles such as plastids. It is preferred to choose a transit peptide that is recognized and compatible with the host in which the transit peptide is employed. In this invention the plastid of choice is the amyloplast.

20 It is preferred that the hybrid polypeptide be located within the amyloplast in cells such as plant cells which synthesize and store starch in amyloplasts. If the host is a bacterial or other cell that does not contain an amyloplast, there need not be a transit peptide coding region.

A terminator is a DNA sequence that terminates the transcription.

X is the coding region for the payload polypeptide, which may be any polypeptide of interest, or chains of amino acids. It may have up to an entire sequence of a known polypeptide or comprise a useful fragment thereof. The payload polypeptide may be a

polypeptide, a fragment thereof, or biologically active protein which is an enzyme, hormone, growth factor, immunoglobulin, dye, etc. Examples of some of the payload polypeptides that can be employed in this invention include, but are not limited to, prolactin (PRL), serum albumin, growth factors and growth hormones, i.e., somatotropin.

5 Serum albumins include bovine, ovine, equine, avian and human serum albumin. Growth factors include epidermal growth factor (EGF), insulin-like growth factor I (IGF-I), insulin-like growth factor II (IGF-II), fibroblast growth factor (FGF), transforming growth factor alpha (TGF-alpha), transforming growth factor beta (TGF-beta), nerve growth factor (NGF), platelet-derived growth factor (PDGF), and recombinant human insulin-like growth
10 factors I (rHuIGF-I) and II (rHuIGF-II). Somatotropins which can be employed to practice this invention include, but are not limited to, bovine, porcine, ovine, equine, avian and human somatotropin. Porcine somatotropin includes delta-7 recombinant porcine somatotropin, as described and claimed in European Patent Application Publication No. 104,920 (Biogen). Preferred payload polypeptides are somatotropin, insulin A and B
15 chains, calcitonin, beta endorphin, urogastrone, beta globin, myoglobin, human growth hormone, angiotensin, proline, proteases, beta-galactosidase, and cellulases.

The hybrid polypeptide, the SER region and the payload polypeptides may also include post-translational modifications known to the art such as glycosylation, acylation, and other modifications not interfering with the desired activity of the polypeptide.

20 **Developing a Hybrid polypeptide**

The SER region is present in genes involved in starch synthesis. Methods for isolating such genes include screening from genomic DNA libraries and from cDNA libraries. Genes can be cut and changed by ligation, mutation agents, digestion, restriction and other such procedures, e.g., as outlined in Maniatis et al., Molecular Cloning, Cold
25 Spring Harbor Labs, Cold Spring Harbor, N.Y. Examples of excellent starting materials for accessing the SER region include, but are not limited to, the following: starch synthases I, II, III, IV, Branching Enzymes I, IIA and B and granule-bound starch synthase (GBSTS). These genes are present in starch-bearing plants such as rice, maize, peas, potatoes, wheat, and the like. Use of a probe of SER made from genomic DNA or cDNA
30 or mRNA or antibodies raised against the SER allows for the isolation and identification

of useful genes for cloning. The starch enzyme-encoding sequences may be modified as long as the modifications do not interfere with the ability of the SER region to encapsulate associated polypeptides.

When genes encoding proteins that are encapsulated into the starch granule are located, then several approaches to isolation of the SER can be employed, as is known to the art. One method is to cut the gene with restriction enzymes at various sites, deleting sections from the N-terminal end and allowing the resultant protein to express. The expressed truncated protein is then run on a starch gel to evaluate the association and dissociation constant of the remaining protein. Marker genes known to the art, e.g., green fluorescent protein gene, may be attached to the truncated protein and used to determine the presence of the marker gene in the starch granule.

Once the SER gene sequence region is isolated it can be used in making the gene fragment sequence that will express the payload polypeptide encapsulated in starch. The SER gene sequence and the gene sequence encoding the payload polypeptide can be ligated together. The resulting fused DNA can then be placed in a number of vector constructs for expression in a number of hosts. The preferred hosts form starch granules in plastids, but the testing of the SER can be readily performed in bacterial hosts such as *E.coli*.

The nucleic acid sequence coding for the payload polypeptide may be derived from DNA, RNA, genomic DNA, cDNA, mRNA or may be synthesized in whole or in part. The sequence of the payload polypeptide can be manipulated to contain mutations such that the protein produced is a novel, mutant protein, so long as biological function is maintained.

When the payload polypeptide-encoding nucleic acid sequence is ligated onto the SER-encoding sequence, the gene sequence for the payload polypeptide is preferably attached at the end of the SER sequence coding for the N-terminus. Although the N-terminus end is preferred, it does not appear critical to the invention whether the payload polypeptide is ligated onto the N-terminus end or the C-terminus end of the SER. Clearly,

the method of forming the recombinant nucleic acid molecules of this invention, whether synthetically, or by cloning and ligation, is not critical to the present invention.

The central region of the hybrid polypeptide is optional. For some applications of the present invention it can be very useful to introduce DNA coding for a convenient
5 protease cleavage site in this region into the recombinant nucleic acid molecule used to express the hybrid polypeptide. Alternatively, it can be useful to introduce DNA coding for an amino acid sequence that is pH-sensitive to form the central region. If the use of the present invention is to develop a pure protein that can be extracted and released from the starch granule by a protease or the like, then a protease cleavage site is useful.

10 Additionally, if the protein is to be digested in an animal then a protease cleavage site may be useful to assist the enzymes in the digestive tract of the animal to release the protein from the starch. In other applications and in many digestive uses the cleavage site would be superfluous.

The central region site may comprise a spacer. A spacer refers to a peptide that
15 joins the proteins comprising a hybrid polypeptide. Usually it does not have any specific activity other than to join the proteins, to preserve some minimum distance, to influence the folding, charge or hydrophobic or hydrophilic nature of the hybrid polypeptide.

Construct Development

Once the ligated DNA which encodes the hybrid polypeptide is formed, then
20 cloning vectors or plasmids are prepared which are capable of transferring the DNA to a host for expressing the hybrid polypeptides. The recombinant nucleic acid sequence of this invention is inserted into a convenient cloning vector or plasmid. For the present invention the preferred host is a starch granule-producing host. However, bacterial hosts can also be employed. Especially useful are bacterial hosts that have been transformed to
25 contain some or all of the starch-synthesizing genes of a plant. The ordinarily skilled person in the art understands that the plasmid is tailored to the host. For example, in a bacterial host transcriptional regulatory promoters include lac, TAC, trp and the like. Additionally, DNA coding for a transit peptide most likely would not be used and a secretory leader that is upstream from the structural gene may be used to get the

polypeptide into the medium. Alternatively, the product is retained in the host and the host is lysed and the product isolated and purified by starch extraction methods or by binding the material to a starch matrix (or a starch-like matrix such as amylose or amylopectin, glycogen or the like) to extract the product.

5 The preferred host is a plant and thus the preferred plasmid is adapted to be useful in a plant. The plasmid should contain a promoter, preferably a promoter adapted to target the expression of the protein in the starch-containing tissue of the plant. The promoter may be specific for various tissues such as seeds, roots, tubers and the like; or, it can be a constitutive promoter for gene expression throughout the tissues of the plant.

10 Well-known promoters include the 10 kD zein (maize) promoter, the CAB promoter, patatin, 35S and 19S cauliflower mosaic virus promoters (very useful in dicots), the polyubiquitin promoter (useful in monocots) and enhancements and modifications thereof known to the art.

15 The cloning vector may contain coding sequences for a transit peptide to direct the plasmid into the correct location. Examples of transit peptide-coding sequences are shown in the sequence tables. Coding sequences for other transit peptides can be used. Transit peptides naturally occurring in the host to be used are preferred. Preferred transit peptide coding regions for maize are shown in the tables and figures hereof. The purpose of the transit peptide is to target the vector to the correct intracellular area.

20 Attached to the transit peptide-encoding sequence is the DNA sequence encoding the N-terminal end of the payload polypeptide. The direction of the sequence encoding the payload polypeptide is varied depending on whether sense or antisense transcription is desired. DNA constructs of this invention specifically described herein have the sequence encoding the payload polypeptide at the N-terminus end but the SER coding region can

25 also be at the N-terminus end and the payload polypeptide sequence following. At the end of the DNA construct is the terminator sequence. Such sequences are well known in the art.

The cloning vector is transformed into a host. Introduction of the cloning vector, preferably a plasmid, into the host can be done by a number of transformation techniques known to the art. These techniques may vary by host but they include microparticle bombardment, micro injection, *Agrobacterium* transformation, "whiskers" technology (U.S. Patent Nos. 5,302,523 and 5,464,765), electroporation and the like. If the host is a plant, the cells can be regenerated to form plants. Methods of regenerating plants are known in the art. Once the host is transformed and the proteins expressed therein, the presence of the DNA encoding the payload polypeptide in the host is confirmable. The presence of expressed proteins may be confirmed by Western Blot or ELISA or as a result of a change in the plant or the cell.

Uses of Encapsulated Protein

There are a number of applications of this invention. The hybrid polypeptide can be cleaved in a pure state from the starch (cleavage sites can be included) and pure protein can be recovered. Alternatively, the encapsulated payload polypeptide within the starch can be used in raw form to deliver protein to various parts of the digestive tract of the consuming animal ("animal" shall include mammals, birds and fish). For example if the starch in which the material is encapsulated is resistant to digestion then the protein will be released slowly into the intestine of the animal, therefore avoiding degradation of the valuable protein in the stomach. Amino acids such as methionine and lysine may be encapsulated to be incorporated directly into the grain that the animal is fed thus eliminating the need for supplementing the diet with these amino acids in other forms.

The present invention allows hormones, enzymes, proteins, proteinaceous nutrients and proteinaceous medicines to be targeted to specific digestive areas in the digestive tracts of animals. Proteins that normally are digested in the upper digestive tract encapsulated in starch are able to pass through the stomach in a nondigested manner and be absorbed intact or in part by the intestine. If capable of passing through the intestinal wall, the payload polypeptides can be used for medicating an animal, or providing hormones such as growth factors, e.g., somatotropin, for vaccination of an animal or for enhancing the nutrients available to an animal.

If the starch used is not resistant to digestion in the stomach (for example the sugary 2 starch is highly digestible), then the added protein can be targeted to be absorbed in the upper digestive tract of the animal. This would require that the host used to produce the modified starch be mutated or transformed to make sugary 2 type starch. The present invention encompasses the use of mutant organisms that form modified starch as hosts. Some examples of these mutant hosts include rice and maize and the like having sugary 1, sugary 2, brittle, shrunken, waxy, amylose extender, dull, opaque, and floury mutations, and the like. These mutant starches and starches from different plant sources have different levels of digestibility. Thus by selection of the host for expression of the DNA and of the animal to which the modified starch is fed, the hybrid polypeptide can be digested where it is targeted. Different proteins are absorbed most efficiently by different parts of the body. By encapsulating the protein in starch that has the selected digestibility, the protein can be supplied anywhere throughout the digestive tract and at specific times during the digestive process.

Another of the advantages of the present invention is the ability to inhibit or express differing levels of glycosylation of the desired polypeptide. The encapsulating procedure may allow the protein to be expressed within the granule in a different glycosylation state than if expressed by other DNA molecules. The glycosylation will depend on the amount of encapsulation, the host employed and the sequence of the polypeptide.

Improved crops having the above-described characteristics may be produced by genetic manipulation of plants known to possess other favorable characteristics. By manipulating the nucleotide sequence of a starch-synthesizing enzyme gene, it is possible to alter the amount of key amino acids, proteins or peptides produced in a plant. One or more genetically engineered gene constructs, which may be of plant, fungal, bacterial or animal origin, may be incorporated into the plant genome by sexual crossing or by transformation. Engineered genes may comprise additional copies of wildtype genes or may encode modified or allelic or alternative enzymes with new properties. Incorporation of such gene construct(s) may have varying effects depending on the amount and type of

gene(s) introduced (in a sense or antisense orientation). It may increase the plant's capacity to produce a specific protein, peptide or provide an improved amino acid balance.

Cloning Enzymes Involved in Starch Biosynthesis

Known cloning techniques may be used to provide the DNA constructs of this invention. The source of the special forms of the SSTS, GBSTS, BE, glycogen synthase (GS), amylopectin, or other genes used herein may be any organism that can make starch or glycogen. Potential donor organisms are screened and identified. Thereafter there can be two approaches: (a) using enzyme purification and antibody/sequence generation following the protocols described herein; (b) using SSTS, GBSTS, BE, GS, amylopectin or other cDNAs as heterologous probes to identify the genomic DNAs for SSTS, GBSTS, BE, GS, amylopectin or other starch-encapsulating enzymes in libraries from the organism concerned. Gene transformation, plant regeneration and testing protocols are known to the art. In this instance it is necessary to make gene constructs for transformation which contain regulatory sequences that ensure expression during starch formation. These regulatory sequences are present in many small grains and in tubers and roots. For example these regulatory sequences are readily available in the maize endosperm in DNA encoding Granule Bound Starch Synthesis (GBSTS), Soluble Starch Synthases (SSTS) or Branching Enzymes (BE) or other maize endosperm starch synthesis pathway enzymes. These regulatory sequences from the endosperm ensure protein expression at the correct developmental time (e.g., ADPG pyrophosphorylase).

In this method we measure starch-binding constants of starch-binding proteins using native protein electrophoresis in the presence of suitable concentrations of carbohydrates such as glycogen or amylopectin. Starch-encapsulating regions can be elucidated using site-directed mutagenesis and other genetic engineering methods known to those skilled in the art. Novel genetically-engineered proteins carrying novel peptides or amino acid combinations can be evaluated using the methods described herein.

EXAMPLES

Example One:

Method for Identification of Starch-encapsulating Proteins

Starch-Granule Protein Isolation:

- 5 Homogenize 12.5 g grain in 25 ml Extraction buffer (50 mM Tris acetate, pH 7.5, 1 mM EDTA, 1 mM DTT for 3 x 20 seconds in Waring blender with 1 min intervals between blending). Keep samples on ice. Filter through mira cloth and centrifuge at 6,000 rpm for 30 min. Discard supernatant and scrape off discolored solids which overlay white starch pellet. Resuspend pellet in 25 ml buffer and recentrifuge. Repeat washes twice
- 10 more. Resuspend washed pellet in -20°C acetone, allow pellet to settle at -20°C. Repeat. Dry starch under stream of air. Store at -20°C.

Protein Extraction:

- Mix 50 mg starch with 1 ml 2% SDS in eppendorf. Vortex, spin at 18,000 rpm, 5 min, 4°C. Pour off supernatant. Repeat twice. Add 1 ml sample buffer (4 ml distilled
- 15 water, 1 ml 0.5 M Tris-HCl, pH 6.8, 0.8 ml glycerol, 1.6 ml 10% SDS, 0.4 ml B-mercaptoethanol, 0.2 ml 0.5% bromphenol blue). Boil eppendorf for 10 min with hole in lid. Cool, centrifuge 10,000 rpm for 10 min. Decant supernatant into new eppendorf. Boil for 4 minutes with standards. Cool.

SDS-Page Gels: (non-denaturing)

20		10% Resolve	4% Stack
	Acryl/Bis 40% stock	2.5 ml	1.0 ml
	1.5 M Tris pH 8.8	2.5 ml	-
	0.5 M Tris pH 8.8	-	2.5 ml
	10% SDS	100 µl	100 µl
25	Water	4.845 ml	6.34 ml
	Degas 15 min add fresh		
	10% Ammonium Persulfate	50 µl	50 µl
	TEMED	5µl	10 µl

Mini-Protean II Dual Slab Cell; 3.5 ml of Resolve buffer per gel. 4% Stack is poured on top. The gel is run at 200V constant voltage. 10 x Running buffer (250 mM Tris, 1.92 M glycine, 1% SDS, pH 8.3).

Method of Measurement of Starch-Encapsulating Regions:

5 Solutions:

	Extraction Buffer:	50 mM Tris-acetate pH 7.5, 10 mM EDTA, 10% sucrose, 2.5 mM DTT-fresh.
	Stacking Buffer:	0.5 M Tris-HCl, pH 6.8
	Resolve Buffer:	1.5 M Tris-HCl, pH 8.8
10	10 X Lower Electrode Buffer:	30.3 g Tris + 144 g Glycine qs to 1 L. (pH is ~8.3, no adjustment). Dilute for use.
	Upper Electrode Buffer:	Same as Lower
	Sucrose Solution:	18.66 g sucrose + 100 ml dH ₂ O
15	30% Acryl/Bis Stock (2.67%C):	146 g acrylamide + 4 g bis + 350 ml dH ₂ O. Bring up to 500 ml. Filter and store at 4 C in the dark for up to 1 month.
	15% Acryl/Bis Stock (20% C):	6 g acrylamide + 1.5 g bis + 25 ml dH ₂ O. Bring up to 50 ml. Filter and store at 4 C in the dark for up to 1 month.
20	Riboflavin Solution:	1.4 g riboflavin + 100 ml dH ₂ O. Store in dark for up to 1 month.
	SS Assay mix:	25 mM Sodium Citrate, 25 mM Bicine-NaOH (pH 8.0), 2 mM EDTA, 1 mM DTT-fresh, 1 mM Adenosine 5' Diphosphoglucose-fresh, 10 mg/ml rabbit
25		liver glycogen Type III-fresh.
	Iodine Solution:	2 g iodine + 20 g KI, 0.1 N HCl up to 1 L.

Extract:

- 4 ml extraction buffer + 12 g endosperm. Homogenize.
- filter through mira cloth or 4 layers cheesecloth, spin 20,000 g (14,500 rpm, SM-24 rotor), 20 min., 4°C.
- 5 • remove supernatant using a glass pipette.
- 0.85 ml extract + 0.1 ml glycerol + 0.05 ml 0.5% bromophenol blue.
- vortex and spin 5 min. full speed microfuge. Use directly or freeze in liquid nitrogen and store at -80°C for up to 2 weeks.

Cast Gels:

- 10 Attach Gel Bond PAG film (FMC Industries, Rockland, ME) to (inside of) outer glass plate using two-sided scotch tape, hydrophilic side up. The tape and the film is lined up as closely and evenly as possible with the bottom of the plate. The film is slightly smaller than the plate. Squirt water between the film and the plate to adhere the film. Use a tissue to push out excess water. Set up plates as usual, then seal the bottom
- 15 of the plates with tacky adhesive. The cassette will fit into the casting stand if the gray rubber is removed from the casting stand. The gel polymerizes with the film, and stays attached during all subsequent manipulations.

Cast 4.5% T resolve mini-gel (0.75 mm):

- 2.25 ml dH₂O
- 20 + 3.75 ml sucrose solution
- + 2.5 ml resolve buffer
- + 1.5 ml 30% Acryl/Bis stock
- + various amounts of glycogen for each gel (i.e., 0 - 1.0%)
- DEGAS 15 MIN.
- 25 + 50 µl 10% APS
- + 5 µl TEMED
- POLYMERIZE FOR 30 MIN. OR OVERNIGHT

Cast 3.125 % T stack:

1.59 ml dH₂O

+ 3.75 ml sucrose solution
+ 2.5 ml stack buffer
+ 2.083 ml 15% Acryl/Bis stock

DO NOT DEGAS

5 15 μ l 10% APS

+ 35 μ l riboflavin solution

+ 30 μ l TEMED

POLYMERIZE FOR 2.5 HOURS CLOSE TO A LIGHT BULB

cool in 4°C before pulling out combs. Can also not use combs, and just

10 cast a centimeter of stacker.

The foregoing procedure:

- Can run at different temperatures; preincubate gels and solutions.
- Pre-run for 15 min. at 200 V
- Load gel: 7 μ l per well, or 115 μ l if no comb.
- 15 · Run at 140 V until dye front is close to bottom. Various running temperatures are achieved by placing the whole gel rig into a water bath. Can occasionally stop the run to insert a temperature probe into the gel.
- Enzyme assay: Cut gels off at dye front. Incubate in SS. Assay mix overnight at room temperature with gentle shaking. Rinse gels with water. Flood with I₂/KI
- 20 solution.
- Take pictures of the gels on a light box, and measure the pictures. R_m = mm from top of gel to the active band/mm from top of gel to the bottom of the gel where it was cut (where the dye front was). Plot % glycogen vs. 1/R_m. The point where the line intersects the x axis is -K (where y=0).

25 **Testing and evaluation protocol for SER region length:**

Following the procedure above for selection of the SER region requires four basic steps. First DNA encoding a protein having a starch-encapsulation region must be selected. This can be selected from known starch-synthesizing genes or starch-binding genes such as genes for amylases, for example. The protein must be extracted. A number

30 of protein extraction techniques are well known in the art. The protein may be treated

with proteases to form protein fragments of different lengths. The preferred fragments have deletions primarily from the N-terminus region of the protein. The SER region is located nearer to the C-terminus end than the N-terminus end. The protein is run on the gels described above and affinity for the gel matrix is evaluated. Higher affinity shows more preference of that region of the protein for the matrix. This method enables comparison of different proteins to identify the starch-encapsulating regions in natural or synthetic proteins.

Example Two:

SER Fusion Vector:

The following fusion vectors are adapted for use in *E.coli*. The fusion gene that was attached to the probable SER in these vectors encoded for the green fluorescent protein (GFP). Any number of different genes encoding for proteins and polypeptides could be ligated into the vectors. A fusion vector was constructed having the SER of waxy maize fused to a second gene or gene fragment, in this case GFP.

pEXS114 (see FIG. 1a): Synthetic GFP (SGFP) was PCR-amplified from the plasmid HBT-SGFP (from Jen Sheen; Dept. of Molecular Biology; Wellman 11, MGH; Boston, MA 02114) using the primers EXS73 (5'-GACTAGTCATATG GTG AGC AAG GGC GAG GAG-3') [SEQ ID NO:1] and EXS74 (5'-CTAGATCTTCATATG CTT GTA CAG CTC GTC CAT GCC-3') [SEQ ID NO:2]. The ends of the PCR product were polished off with T DNA polymerase to generate blunt ends; then the PCR product was digested with *Spe* I. This SGFP fragment was subcloned into the *EcoRV-Spe* I sites of pBSK (Stratagene at 11011 North Torrey Pines Rd. La Jolla, Ca.) to generate pEXS114.

pEXS115 [see FIG. 1b]: Synthetic GFP (SGFP) was PCR-amplified from the plasmid HBT-SGFP (from Jen Sheen) using the primers EXS73 (see above) and EXS75 (5'-CTAGATCTTGGCCATGGC CTT GTA CAG CTC GTC CAT GCC-3') [SEQ ID NO:3]. The ends of the PCR product were polished off with T DNA polymerase to generate blunt ends; then the PCR product was digested with *Spe* I. This SGFP fragment was subcloned into the *EcoRV-Spe* I sites of pBSK (Stratagene) generating pEXS115.

pEXSWX (see FIG. 2a): Maize WX subcloned *NdeI-Not I* into pET-21a (see FIG. 2b). The genomic DNA sequence and associated amino acids from which the mRNA sequence can be generated is shown in TABLES 1a and 1b below and alternatively the DNA listed in the following tables could be employed.

TABLE 1a
DNA Sequence and Deduced Amino Acid Sequence
of the waxy Gene in Maize
[SEQ ID NO:4 and SEQ ID NO:5]

	LOCUS	ZMWAXY	4800 bp	DNA	PLN
10	DEFINITION	Zea mays waxy (wx+) locus for UDP-glucose starch glycosyl transferase.			
	ACCESSION	X03935 M24258			
	KEYWORDS	glycosyl transferase; transit peptide; UDP-glucose starch glycosyl transferase; waxy locus.			
15	SOURCE	maize.			
	ORGANISM	Zea mays Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida; Commelinidae; Cyperales; Poaceae.			
	REFERENCE	1 (bases 1 to 4800)			
20	AUTHORS	Kloesgen, R.B., Gierl, A., Schwarz-Sommer, Z. and Saedler, H.			
	TITLE	Molecular analysis of the waxy locus of Zea mays			
	JOURNAL	Mol. Gen. Genet. 203, 237-244 (1986)			
	STANDARD	full automatic			
	COMMENT	NCBI gi: 22509			
25	FEATURES	Location/Qualifiers			
	source	1..4800			
	repeat_region	283..287 /organism="Zea mays"			
30	repeat_region	288..292 /note="direct repeat 1"			
	repeat_region	293..297 /note="direct repeat 1"			
35	repeat_region	298..302 /note="direct repeat 1"			
	misc_feature	372..385 /note="GC stretch (pot. regulatory factor binding site)"			
40	misc_feature	442..468 /note="GC stretch (pot. regulatory factor binding site)"			
	misc_feature	768..782 /note="GC stretch (pot. regulatory factor binding site)"			
45	misc_feature	810..822 /note="GC stretch (pot. regulatory factor binding site)"			
	misc_feature	821..828 /note="target duplication site (Ac7)"			
50	CAAT_signal	821..828			
	TATA_signal	867..873			
	misc_feature	887..900 /note="GC stretch (pot. regulatory factor binding site)"			
55	misc_feature	901 /note="transcriptional start site"			
	exon	901..1080 /number=1			

```

intron      1081..1219
            /number=1
exon        1220..1553
            /number=2
5 transit_peptide 1233..1448
CDS         join(1449..1553,1685..1765,1860..1958,2055..2144,
2226..2289,2413..2513,2651..2760,2858..3101,3212..3394,
10          3490..3681,3793..3879,3977..4105,4227..4343)
            /note="NCBI gi: 22510"
            /codon_start=1
            /product="glucosyl transferase"

15 /translation="ASAGMNVVFGAEMAPWSKTGGLGDLGGLPPAMAANGHRVMV
SPRYDQYKDAWDTSVVSEIKMGDGYETVRFFHCYKRGVDRVFDHPLFLERVWGKTEE
KIYGPVAGTDYRDNQLRFSLLCQAALAPRILSLNPNPYFSGPYGEDVVFVCNDWHTG
20 PLSCYLKSNYQSHGIYRDAKTAFCIHNISYQGRFAFSDYPELNLPERFKSSFDFIDGY
EKPVEGRKINWMKAGILEADRVLTVPYYAEELISGIARGCELDNIMRLTGITGIVNG
MDVSEWDPSRDKYIAVKYDVSTAVEAKALNKEALQAEVGLPVDRNIPLVAFIGRLEEQ
25 KGPDMAAAIPQLMEMVEDVQIVLLGTGKKKFERMLMSAEKFPKGKVRVVKFNAALA
HHIMAGADVLAVTSRFEPGLIQLQGMRYGTPCACASTGGLVDTIIEGKTGFHMGRLS
30 VDCNVVEPADVKKVATTLQRAIKVVGTPAYEEMVRNCMIQDLSWKGPKNWENVLLSL
   GVAGGEPGVEGEEIAPLAKENVAAP"
intron      1554..1684
            /number=2
35 exon      1685..1765
            /number=3
intron      1766..1859
            /number=3
exon        1860..1958
            /number=4
40 intron    1959..2054
            /number=4
exon        2055..2144
            /number=5
45 intron    2145..2225
            /number=5
exon        2226..2289
            /number=6
intron      2290..2412
            /number=6
50 exon      2413..2513
            /number=7
intron      2514..2650
            /number=7
55 exon      2651..2760
            /number=8
intron      2761..2857
            /number=8
exon        2858..3101
            /number=9
60 intron    3102..3211
            /number=9
exon        3212..3394
            /number=10
65 misc_feature 3358..3365
            /note="target duplication site (Ac9)"
intron      3395..3489
            /number=10
exon        3490..3681

```

```

misc_feature      /number=11
                  3570..3572
                  /note="target duplication site (Spm 18)"
5      intron      3682..3792
                  /number=11
      exon         3793..3879
                  /number=12
      intron      3880..3976
                  /number=12
10     exon        3977..4105
                  /number=13
      intron      4106..4226
                  /number=13
      exon        4227..4595
                  /number=14
15     polyA_signal 4570..4575
      polyA_signal 4593..4598
      polyA_site   4595
      polyA_signal 4597..4602
20     polyA_site   4618
      polyA_site   4625
BASE COUNT      935 A   1413 C   1447 G   1005 T
ORIGIN
25      1 CAGCGACCTA TTACACAGCC CGCTCGGGCC CGCGACGTCG GGACACATCT TCTTCCCCCT
      61 TTTGGTGAAG CTCTGCTCGC AGCTGTCCGG CTCCTTGGAC GTTCGTGTGG CAGATTCATC
      121 TGTTGTCTCG TCTCCTGTGC TTCCTGGGTA GCTTGTGTAG TGGAGCTGAC ATGGTCTGAG
30     181 CAGGCTTAAA ATTTGCTCGT AGACGAGGAG TACCAGCACA GCACGTTGCG GATTTCTCTG
      241 CCTGTGAAGT GCAACGTCTA GGATTGTACAC ACGCCTTGGT CGCGTCGCGT CGCGTCGCGT
      301 CGATGCGGTG GTGAGCAGAG CAGCAACAGC TGGGCGGCCC AACGTTGGCT TCCGTGTCTT
35     361 CGTCGTACGT ACGCGCGCGC CGGGGACACG CAGCAGAGAG CGGAGAGCGA GCCGTGCACG
      421 GGGAGGTGGT GTGGAAGTGG AGCCGCGCGC CCGGCCGCCC GCGCCCGGTG GGCAACCCAA
40     481 AAGTACCCAC GACAAGCGAA GGCGCCAAAG CGATCCAAGC TCCGGAACGC AACAGCATGC
      541 GTCGCGTCGG AGAGCCAGCC ACAAGCAGCC GAGAACCGAA CCGGTGGGCG ACGCGTCATG
      601 GGACGGACGC GGGCGACGCT TCCAAACGGG CCACGTACGC CGGCGTGTGC GTGCGTGCAG
45     661 ACGACAAGCC AAGGCGAGGC AGCCCCGAT CGGGAAAGCG TTTTGGGCGC GAGCGCTGGC
      721 GTGCGGGTCA GTCGCTGGTG CGCAGTGCCG GGGGGAACGG GTATCGTGGG GGGCGCGGGC
50     781 GGAGGAGAGC GTGGCGAGGG CCGAGAGCAG CGCGCGGCCG GGTACGCAA CGCGCCCCAC
      841 GTACTGCCCT CCCCCTCCGC GCGCGCTAGA AATACCGAGG CCTGGACCGG GGGGGGGCCC
      901 CGTCACATCC ATCCATCGAC CGATCGATCG CCACAGCCAA CACCACCCGC CGAGGCGACG
55     961 CGACAGCCGC CAGGAGGAAG GAATAAACTC ACTGCCAGCC AGTGAAGGGG GAGAAGTGTA
      1021 CTGCTCCGTC GACCACTGCG CGCACCGCCC GGCAGGGCTG CTCATCTCGT CGACGACCAG
60     1081 GTTCTGTTCC GTTCCGATCC GATCCGATCC TGTCCTTGAG TTTCGTCCAG ATCCTGGCGC
      1141 GTATCTGCGT GTTTGATGAT CCAGGTTCTT CGAACCTAAA TCTGTCCGTG CACACGTCTT
      1201 TTCTCTCTCT CCTACGCAGT GGATTAATCG GCATGGCGGC TCTGGCCACG TCGCAGCTCG
65     1261 TCGCAACGCG CGCCGGCCTG GCGGTCCCGG ACGCGTCCAC GTTCCGCCGC GGCGCCGCGC

```

1321 AGGGCCTGAG GGGGGCCCGG GCGTCGGCGG CGGCGGACAC GCTCAGCATG CGGACCAGCG
 1381 CGCGCGCGGC GCCCAGGCAC CAGCAGCAGG CGCGCCGCGG GGGCAGGTTT CCGTCGCTCG
 5 1441 TCGTGTGCGC CAGCGCCGGC ATGAACGTCG TCTTCGTCGG CGCCGAGATG GCGCCGTGGA
 1501 GCAAGACCGG CGGCCTCGGC GACGTCCTCG GCGGCCTGCC GCCGGCCATG GCCGTAAGCG
 10 1561 CGCGCACCGA GACATGCATC CGTTGGATCG CGTCTTCTTC GTGCTCTTGC CGCGTGCATG
 1621 ATGCATGTGT TTCCTCCTGG CTTGTGTTTC TGTATGTGAC GTGTTTGTTT GGGCATGCAT
 1681 GCAGGCGAAC GGGCACCGTG TCATGGTCGT CTCTCCCCGC TACGACCAGT ACAAGGACGC
 15 1741 CTGGGACACC AGCGTCGTGT CCGAGGTACG GCCACCGAGA CCAGATTGAG ATCACAGTCA
 1801 CACACACCGT CATATGAACC TTTCTCTGCT CTGATGCCTG CAACTGCAAA TGCATGCAGA
 1861 TCAAGATGGG AGACGGGTAC GAGACGGTCA GGTTCCTTCA CTGCTACAAG CGCGGAGTGG
 20 1921 ACCGCGTGTT CGTTGACCAC CCACTGTTCC TGGAGAGGGT GAGACGAGAT CTGATCACTC
 1981 GATACGCAAT TACCACCCCA TTGTAAGCAG TTACAGTGAG CTTTTTTTCC CCCCAGCCTG
 25 2041 GTCGCTGGTT TCAGGTTTGG GGAAAGACCG AGGAGAAGAT CTACGGGCCT GTCGCTGGAA
 2101 CGGACTACAG GGACAACCAG CTGCGGTTCA GCCTGCTATG CCAGGTCAGG ATGGCTTGGT
 2161 ACTACAACTT CATATCATCT GTATGCAGCA GTATACTG ATGAGAAATG CATGCTGTTC
 30 2221 TGCAGGCAGC ACTTGAAGCT CCAAGGATCC TGAGCCTCAA CAACAACCCA TACTTCTCCG
 2281 GACCATACGG TAAGAGTTGC AGTCTTCGTA TATATATCTG TTGAGCTCGA GAATCTTCAC
 35 2341 AGGAAGCGGC CCATCAGACG GACTGTCATT TTACACTGAC TACTGCTGCT GCTCTTCGTC
 2401 CATCCATACA AGGGGAGGAC GTCGTGTTTC TCTGCAACGA CTGGCACACC GGCCCTCTCT
 2461 CGTGCTACCT CAAGAGCAAC TACCAGTCCC ACGGCATCTA CAGGGACGCA AAGGTTGCCT
 40 2521 TCTCTGAACT GAACAACGCC GTTTTCGTTT TCCATGCTCG TATATACCTC GTCTGGTAGT
 2581 GGTGGTGCTT CTCTGAGAAA CTAAGTAAA CTGACTGCAT GTCTGTCTGA CCATCTTCAC
 45 2641 GTACTACCAG ACCGCTTTCT GCATCCACAA CATCTCCTAC CAGGGCCGGT TCGCCTTCTC
 2701 CGACTACCCG GAGCTGAACC TCCCGGAGAG ATTCAAGTCG TCCTTCGATT TCATCGACGG
 2761 GTCTGTTTTT CTGCGTGCAT GTGAACATTC ATGAATGGTA ACCCACAACCT GTTCGCGTCC
 50 2821 TGCTGGTTCA TTATCTGACC TGATTGCATT ATTGCAGCTA CGAGAAGCCC GTGGAAGGCC
 2881 GGAAGATCAA CTGGATGAAG GCCGGGATCC TCGAGGCCGA CAGGGTCCTC ACCGTCAGCC
 55 2941 CCTACTACGC CGAGGAGCTC ATCTCCGGCA TCGCCAGGGG CTGCGAGCTC GACAACATCA
 3001 TGCGCCTCAC CGGCATCACC GGCATCGTCA ACGGCATGGA CGTCAGCGAG TGGGACCCCA
 3061 GCAGGGACAA GTACATCGCC GTGAAGTACG ACGTGTGACG GGTGAGCTGG CTAGCTCTGA
 60 3121 TTCTGCTGCC TGGTCCTCCT GCTCATCATG CTGGTTCGGT ACTGACGCGG CAAGTGTACG
 3181 TACGTGCGTG CGACGGTGGT GTCCGGTTCA GGCCGTGGAG GCCAAGGCGC TGAACAAGGA
 65 3241 GCGGCTGCAG GCGGAGGTCG GGCTCCCGGT GGACCGGAAC ATCCCGCTGG TGGCGTTCAT
 3301 CGGCAGGCTG GAAGAGCAGA AGGGCCCCGA CGTCATGGCG GCCGCCATCC CGCAGCTCAT

	3361	GGAGATGGTG	GAGGACGTGC	AGATCGTTCT	GCTGGTACGT	GTGCGCCGGC	CGCCACCCGG
	3421	CTACTACATG	CGTGTATCGT	TCGTTCTACT	GGAACATGCC	TGTGAGCAAC	GCGATGGATA
5	3481	ATGCTGCAGG	GCACGGGCAA	GAAGAAGTTC	GAGCGCATGC	TCATGAGCGC	CGAGGAGAAG
	3541	TTCCCAGGCA	AGGTGCGCGC	CGTGGTCAAG	TTCAACGCGG	CGCTGGCGCA	CCACATCATG
10	3601	GCCGGCGCCG	ACGTGCTCGC	CGTCACCAGC	CGCTTCGAGC	CCTGCGGCCT	CATCCAGCTG
	3661	CAGGGGATGC	GATACGGAAC	GGTACGAGAG	AAAAAAAAAA	TCCTGAATCC	TGACGAGAGG
	3721	GACAGAGACA	GATTATGAAT	GCTTCATCGA	TTTGAATTGA	TTGATCGATG	TCTCCCGCTG
15	3781	CGACTCTTGC	AGCCCTGCGC	CTGCGCGTCC	ACCGGTGGAC	TCGTCGACAC	CATCATCGAA
	3841	GGCAAGACCG	GGTTCACAT	GGGCCGCCTC	AGCGTCGACG	TAAGCCTAGC	TCTGCCATGT
20	3901	TCTTTCTTCT	TTCTTTCTGT	ATGTATGTAT	GAATCAGCAC	CGCCGTTCTT	GTTTCGTCGT
	3961	CGTCCTCTCT	TCCCAGTGTA	ACGTCGTGGA	GCCGGCGGAC	GTCAAGAAGG	TGGCCACCAC
	4021	ATTGCAGCGC	GCCATCAAGG	TGGTCGGCAC	GCCGGCGTAC	GAGGAGATGG	TGAGGAACTG
25	4081	CATGATCCAG	GATCTCTCCT	GGAAGGTACG	TACGCCCCGC	CCGCCCCGCC	CCGCCAGAGC
	4141	AGAGCGCCAA	GATCGACCGA	TCGACCGACC	ACACGTACGC	GCCTCGCTCC	TGTCGCTGAC
30	4201	CGTGGTTTAA	TTTGCGAAAT	GCGCAGGGCC	CTGCCAAGAA	CTGGGAGAAC	GTGCTGCTCA
	4261	GCCTCGGGGT	CGCCGGCGGC	GAGCCAGGGG	TCGAAGGCCA	GGAGATCGCG	CCGCTCGCCA
	4321	AGGAGAACGT	GGCCGCGCCC	TGAAGAGTTC	GGCCTGCAGG	GCCCCTGATC	TCGCGCGTGG
35	4381	TGCAAAGATG	TTGGGACATC	TTCTTATATA	TGCTGTTTCG	TTTATGTGAT	ATGGACAAGT
	4441	ATGTGTAGCT	GCTTGCTTGT	GCTAGTGTA	TGTAGTGTAG	TGGTGGCCAG	TGGCACAACC
40	4501	TAATAAGCGC	ATGAACTAAT	TGCTTGCGTG	TGTAGTTAAG	TACCGATCGG	TAATTTTATA
	4561	TTGCGAGTAA	ATAAATGGAC	CTGTAGTGGT	GGAGTAAATA	ATCCCTGCTG	TTCGGTGTTC
	4621	TTATCGCTCC	TCGTATAGAT	ATTATATAGA	GTACATTTTT	CTCTCTCTGA	ATCCTACGTT
45	4681	TGTGAAATTT	CTATATCATT	ACTGTAAAAT	TTCTGCGTTC	CAAAAGAGAC	CATAGCCTAT
	4741	CTTTGGCCCT	GTTTGTTTCG	GCTTCTGGCA	GCTTCTGGCC	ACCAAAGCT	GCTGCGGACT

//

TABLE 1b
DNA Sequence and Deduced Amino Acid Sequence in waxy Gene in Rice
[SEQ ID NO:6 and SEQ ID NO:7]

```

5  LOCUS       OSWX             2542 bp      RNA             PLN
   DEFINITION  O.sativa Waxy mRNA.
   ACCESSION   X62134 S39554
   KEYWORDS    glucosyltransferase; starch biosynthesis; waxy gene.
   SOURCE      rice.
10  ORGANISM   Oryza sativa
                Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida;
                Commelinidae; Cyperales; Poaceae.
   REFERENCE   1 (bases 1 to 2542)
   AUTHORS     Okayaki,R.J.
15  TITLE      Direct Submission
   JOURNAL      Submitted (12-SEP-1991) to the EMBL/GenBank/DDBJ databases.
                R.J.
                Okayaki, University of Florida, Dep of Vegetable Crops, 1255
                Fifield Hall, 514 IFAS, Gainesville, Florida 32611-0514, USA
20  STANDARD   full automatic
   REFERENCE   2 (bases 1 to 2542)
   AUTHORS     Okagaki,R.J.
   TITLE       Nucleotide sequence of a long cDNA from the rice waxy gene
   JOURNAL      Plant Mol. Biol. 19, 513-516 (1992)
25  STANDARD   full automatic
   COMMENT     NCBI gi: 20402
   FEATURES     Location/Qualifiers
                source                1..2542
                /organism="Oryza sativa"
                /dev_stage="immature seed"
                /tissue_type="seed"
                CDS                    453..2282
                /gene="Wx"
                /standard_name="Waxy gene"
35  /EC_number="2.4.1.21"
                /note="NCBI gi: 20403"
                /codon_start=1
                /function="starch biosynthesis"
                /product="starch (bacterial glycogen) synthase"
40  /translation="MSALTTSQLATSATGFGIADRSAPSSLLRHGFQGLKPRSPAGGD
                ATLSLVTT SARATPKQQRSVQRGSRFPVSVVYATGAGMNVFVGAEMAPWSKTGGLG
                DVLGGLPPAMAANGHRVMVISPRYDQYKDAWDTSVVAEIKVADRYERVRFFHCYKRGV
                DRVFIDHPSFLEKVGKTGEKIYGPDTGVDYKDNQMRFSLLCQAALAPRILNLNNNP
                YFKGTYGEDVVFVCNDWHTGPLASYLKNNYQPNGIYRNAKVAFCIHNISYQGRFAFED
50  YPELNLSEFRSSFDIDGYDTPVEGRKINWMKAGILEADRVLTVSPYYAEELISGIA
                RGCELDNIMRLTGITGIVNGMDVSEWDPSKDKYITAKYDATTAEAKALNKEALQAEA
                GLPVDRKIPLIAFIGRL EEQKGPDMVMAAAIPELMQEDVQIVLLGTGKKKFELLLKSME
                EKYPGKVRVAVVKFNAPLAHLIMAGADVLA VPSRFEPGLIQLQGMRYGTPCACASTGG
                LVDTVIEGKTGFHMGRLSVDCKVVEPSDVKKVAATLKRAIKVVGTPAYEEMVRNCMNQ
60  DLSWKGP AKNWENVLLGLGVAGSAPGIEGDEIAPLAKENVAAP"
                3'UTR                    2283..2535
                polyA_site            2535
   BASE COUNT  610 A      665 C      693 G      574 T
   ORIGIN
65  1 GAATTCAGTG TGAAGGAATA GATTCTCTC AAAACAATTT AATCATTCAT CTGATCTGCT

```


61 CAAAGCTCTG TGCATCTCCG GGTGCAACGG CCAGGATATT TATTGTGCAG TAAAAAATG
 121 TCATATCCCC TAGCCACCCA AGAACTGCT CCTTAAGTCC TTATAAGCAC ATATGGCATT
 5 181 GTAATATATA TGTTTGAGTT TTAGCGACAA TTTTTTTAAA AACTTTTGGT CCTTTTTATG
 241 AACGTTTTTAA GTTTCACGTG CTTTTTTTTT CGAATTTTAA ATGTAGCTTC AAATTCTAAT
 301 CCCCAATCCA AATTGTAATA AACTTCAATT CTCCTAATTA ACATCTTAAT TCATTTATTT
 10 361 GAAAACCAGT TCAAATTCTT TTTAGGCTCA CCAAACCTTA AACAATTCAA TTCAGTGCAG
 421 AGATCTTCCA CAGCAACAGC TAGACAACCA CCATGTCGGC TCTCACCACG TCCCAGCTCG
 15 481 CCACCTCGGC CACCGGCTTC GGCATCGCCG ACAGGTCGGC GCCGTCTGTC CTGCTCCGCC
 541 ACGGGTTCCA GGGCCTCAAG CCCCGCAGCC CCGCCGGCGG CGACGCGACG TCGCTCAGCG
 601 TGACGACCAG CGCGCGCGCG ACGCCCAAGC AGCAGCGGTC GGTGCAGCGT GGCAGCCGGA
 20 661 GGTTCCCCTC CGTCGTCGTG TACGCCACCG GCGCCGGCAT GAACGTCGTG TTCGTCGGCG
 721 CCGAGATGGC CCCCTGGAGC AAGACCGGCG GCCTCGGTGA CGTCCTCGGT GGCCTCCCCC
 25 781 CTGCCATGGC TGCGAATGGC CACAGGTGCA TGGTGATCTC TCCTCGGTAC GACCAGTACA
 841 AGGACGCTTG GGATACCAGC GTTGTGGCTG AGATCAAGGT TGCAGACAGG TACGAGAGGG
 901 TGAGGTTTTT CCATTGCTAC AAGCGTGGAG TCGACCGTGT GTTCATCGAC CATCCGTCAT
 30 961 TCCTGGAGAA GGTTTGGGGA AAGACCGGTG AGAAGATCTA CGGACCTGAC ACTGGAGTTG
 1021 ATTACAAAGA CAACCAGATG CGTTTCAGCC TTCTTTGCCA GGCAGCACTC GAGGCTCCTA
 35 1081 GGATCCTAAA CCTCAACAAC AACCATACT TCAAAGGAAC TTATGGTGAG GATGTTGTGT
 1141 TCGTCTGCAA CGACTGGCAC ACTGGCCAC TGGCGAGCTA CCTGAAGAAC AACTACCAGC
 1201 CCAATGGCAT CTACAGGAAT GCAAAGGTTG CTTTCTGCAT CCACAACATC TCCTACCAGG
 40 1261 GCCGTTTCGC TTTTCAGGAT TACCCTGAGC TGAACCTCTC CGAGAGGTTT AGGTCATCCT
 1321 TCGATTTTCAT CGACGGGTAT GACACGCCGG TGGAGGGCAG GAAGATCAAC TGGATGAAGG
 45 1381 CCGGAATCCT GGAAGCCGAC AGGGTGCTCA CCGTGAGCCC GTACTACGCC GAGGAGCTCA
 1441 TCTCCGGCAT CGCCAGGGGA TGCGAGCTCG ACAACATCAT GCGGCTCACC GGCATCACCG
 1501 GCATCGTCAA CGGCATGGAC GTCAGCGAGT GGGATCCTAG CAAGGACAAG TACATCACCG
 50 1561 CCAAGTACGA CGCAACCAGC GCAATCGAGG CGAAGGCGCT GAACAAGGAG GCGTTGCAGG
 1621 CGGAGGCGGG TCTTCCGGTC GACAGGAAAA TCCCACTGAT CGCGTTCATC GGCAGGCTGG
 55 1681 AGGAACAGAA GGGCCCTGAC GTCATGGCCG CCGCCATCCC GGAGCTCATG CAGGAGGACG
 1741 TCCAGATCGT TCTTCTGGGT ACTGGAAAGA AGAAGTTCGA GAAGCTGCTC AAGAGCATGG
 1801 AGGAGAAGTA TCCGGGCAAG GTGAGGGCGG TGGTGAAGTT CAACGCGCCG CTTGCTCATC
 60 1861 TCATCATGGC CGGAGCCGAC GTGCTCGCCG TCCCAGCCG CTTGAGCCC TGTGGACTCA
 1921 TCCAGCTGCA GGGGATGAGA TACGGAACGC CCTGTGCTTG CGCGTCCACC GGTGGGCTCG
 65 1981 TGGACACGGT CATCGAAGGC AAGACTGGTT TCCACATGGG CCGTCTCAGC GTCGACTGCA
 2041 AGGTGGTGGA GCCAAGCGAC GTGAAGAAGG TGGCGGCCAC CCTGAAGCGC GCCATCAAGG

2101 TCGTCGGCAC GCCGGCGTAC GAGGAGATGG TCAGGAACCTG CATGAACCAG GACCTCTCCT
 2161 GGAAGGGGCC TGCGAAGAAC TGGGAGAATG TGCTCCTGGG CCTGGGCGTC GCCGGCAGCG
 5 2221 CGCCGGGGAT CGAAGGCGAC GAGATCGCGC CGCTCGCCAA GGAGAACGTG GCTGCTCCTT
 2281 GAAGAGCCTG AGATCTACAT ATGGAGTGAT TAATTAATAT AGCAGTATAT GGATGAGAGA
 10 2341 CGAATGAACC AGTGGTTTGT TTGTTGTAGT GAATTTGTAG CTATAGCCAA TTATATAGGC
 2401 TAATAAGTTT GATGTTGTAC TCTTCTGGGT GTGCTTAAGT ATCTTATCGG ACCCTGAATT
 2461 TATGTGTGTG GCTTATTGCC AATAATATTA AGTAATAAAG GGTATTATTAT ATTATTATAT
 15 2521 ATGTTATATT ATACTAAAAA AA

//

TABLE 2
DNA Sequence and Deduced Amino Acid Sequence of
the Soluble Starch Synthase IIa Gene in Maize
[SEQ ID NO:8 and SEQ ID NO:9]

FILE NAME : MSS2C.SEQ SEQUENCE : NORMAL 2007 BP

CODON TABLE : UNIV.TCN

SEQUENCE REGION : 1 - 2007

TRANSLATION REGION : 1 - 2007

*** DNA TRANSLATION ***

1	GCT GAG GCT GAG GCC GGG GGC AAG GAC GCG CCG CCG GAG AGG AGC GGC	48
1	A E A E A G G K D A P P E R S G	16
49	GAC GCC GCC AGG TTG CCC CGC GCT CGG CGC AAT GCG GTC TCC AAA CGG	96
30 17	D A A R L P R A R R N A V S K R	32
97	AGG GAT CCT CTT CAG CCG GTC GGC CGG TAC GGC TCC GCG ACG GGA AAC	144
33	R D P L Q P V G R Y G S A T G N	48
145	ACG GCC AGG ACC GGC GCC GCG TCC TGC CAG AAC GCC GCA TTG GCG GAC	192
49	T A R T G A A S C Q N A A L A D	64
35 193	GTT GAG ATC GTT GAG ATC AAG TCC ATC GTC GCC GCG CCG CCG ACG AGC	240
65	V E I V E I K S I V A A P P T S	80
241	ATA GTG AAG TTC CCA GGG CGC GGG CTA CAG GAT GAT CCT TCC CTC TGG	288
81	I V K F P G R G L Q D D P S L W	96
289	GAC ATA GCA CCG GAG ACT GTC CTC CCA GCC CCG AAG CCA CTG CAT GAA	336
40 97	D I A P E T V L P A P K P L H E	112
337	TCG CCT GCG GTT GAC GGA GAT TCA AAT GGA ATT GCA CCT CCT ACA GTT	384
113	S P A V D G D S N G I A P P T V	128
385	GAG CCA TTA GTA CAG GAG GCC ACT TGG GAT TTC AAG AAA TAC ATC GGT	432
129	E P L V Q E A T W D F K K Y I G	144
45 433	TTT GAC GAG CCT GAC GAA GCG AAG GAT GAT TCC AGG GTT GGT GCA GAT	480

	145	F	D	E	P	D	E	A	K	D	D	S	R	V	G	A	D	160
	481	GAT	GCT	GGT	TCT	TTT	GAA	CAT	TAT	GGG	ACA	ATG	ATT	CTG	GGC	CTT	TGT	528
	161	D	A	G	S	F	E	H	Y	G	T	M	I	L	G	L	C	176
5	529	GGG	GAG	AAT	GTT	ATG	AAC	GTG	ATC	GTG	GTG	GCT	GCT	GAA	TGT	TCT	CCA	576
	177	G	E	N	V	M	N	V	I	V	V	A	A	E	C	S	P	192
	577	TGG	TGC	AAA	ACA	GGT	GGT	CTT	GGA	GAT	GTT	GTG	GGA	GCT	TTA	CCC	AAG	624
	193	W	C	K	T	G	G	L	G	D	V	V	G	A	L	P	K	208
	625	GCT	TTA	GCG	AGA	AGA	GGA	CAT	CGT	GTT	ATG	GTT	GTG	GTA	CCA	AGG	TAT	672
	209	A	L	A	R	R	G	H	R	V	M	V	V	V	P	R	Y	224
10	673	GGG	GAC	TAT	GTG	GAA	GCC	TTT	GAT	ATG	GGA	ATC	CGG	AAA	TAC	TAC	AAA	720
	225	G	D	Y	V	E	A	F	D	M	G	I	R	K	Y	Y	K	240
	721	GCT	GCA	GGA	CAG	GAC	CTA	GAA	GTG	AAC	TAT	TTC	CAT	GCA	TTT	ATT	GAT	768
	241	A	A	G	Q	D	L	E	V	N	Y	F	H	A	F	I	D	256
15	769	GGA	GTC	GAC	TTT	GTG	TTC	ATT	GAT	GCC	TCT	TTC	CGG	CAC	CGT	CAA	GAT	816
	257	G	V	D	F	V	F	I	D	A	S	F	R	H	R	Q	D	272
	817	GAC	ATA	TAT	GGG	GGA	AGT	AGG	CAG	GAA	ATC	ATG	AAG	CGC	ATG	ATT	TTG	864
	273	D	I	Y	G	G	S	R	Q	E	I	M	K	R	M	I	L	288
	865	TTT	TGC	AAG	GTT	GCT	GTT	GAG	GTT	CCT	TGG	CAC	GTT	CCA	TGC	GGT	GGT	912
	289	F	C	K	V	A	V	E	V	P	W	H	V	P	C	G	G	304
20	913	GTG	TGC	TAC	GGA	GAT	GGA	AAT	TTG	GTG	TTC	ATT	GCC	ATG	AAT	TGG	CAC	960
	305	V	C	Y	G	D	G	N	L	V	F	I	A	M	N	W	H	320
	961	ACT	GCA	CTC	CTG	CCT	GTT	TAT	CTG	AAG	GCA	TAT	TAC	AGA	GAC	CAT	GGG	1008
	321	T	A	L	L	P	V	Y	L	K	A	Y	Y	R	D	H	G	336
25	1009	TTA	ATG	CAG	TAC	ACT	CGC	TCC	GTC	CTC	GTC	ATA	CAT	AAC	ATC	GGC	CAC	1056
	337	L	M	Q	Y	T	R	S	V	L	V	I	H	N	I	G	H	352
	1057	CAG	GGC	CGT	GGT	CCT	GTA	CAT	GAA	TTC	CCG	TAC	ATG	GAC	TTG	CTG	AAC	1104
	353	Q	G	R	G	P	V	H	E	F	P	Y	M	D	L	L	N	368
	1105	ACT	AAC	CTT	CAA	CAT	TTC	GAG	CTG	TAC	GAT	CCC	GTC	GGT	GGC	GAG	CAC	1152
	369	T	N	L	Q	H	F	E	L	Y	D	P	V	G	G	E	H	384
30	1153	GCC	AAC	ATC	TTT	GCC	GCG	TGT	GTT	CTG	AAG	ATG	GCA	GAC	CGG	GTG	GTG	1200
	385	A	N	I	F	A	A	C	V	L	K	M	A	D	R	V	V	400
	1201	ACT	GTC	AGC	CGC	GGC	TAC	CTG	TGG	GAG	CTG	AAG	ACA	GTG	GAA	GGC	GGC	1248
	401	T	V	S	R	G	Y	L	W	E	L	K	T	V	E	G	G	416
35	1249	TGG	GGC	CTC	CAC	GAC	ATC	ATC	CGT	TCT	AAC	GAC	TGG	AAG	ATC	AAT	GGC	1296
	417	W	G	L	H	D	I	I	R	S	N	D	W	K	I	N	G	432
	1297	ATT	CGT	GAA	CGC	ATC	GAC	CAC	CAG	GAG	TGG	AAC	CCC	AAG	GTG	GAC	GTG	1344
	433	I	R	E	R	I	D	H	Q	E	W	N	P	K	V	D	V	448
	1345	CAC	CTG	CGG	TCG	GAC	GGC	TAC	ACC	AAC	TAC	TCC	CTC	GAG	ACA	CTC	GAC	1392
	449	H	L	R	S	D	G	Y	T	N	Y	S	L	E	T	L	D	464
40	1393	GCT	GGA	AAG	CGG	CAG	TGC	AAG	GCG	GCC	CTG	CAG	CGG	GAC	GTG	GGC	CTG	1440
	465	A	G	K	R	Q	C	K	A	A	L	Q	R	D	V	G	L	480
	1441	GAA	GTG	CGC	GAC	GAC	GTG	CCG	CTG	CTC	GGC	TTC	ATC	GGG	CGT	CTG	GAT	1488
	481	E	V	R	D	D	V	P	L	L	G	F	I	G	R	L	D	496
45	1489	GGA	CAG	AAG	GGC	GTG	GAC	ATC	ATC	GGG	GAC	GCG	ATG	CCG	TGG	ATC	GCG	1536
	497	G	Q	K	G	V	D	I	I	G	D	A	M	P	W	I	A	512

5	1537	GGG CAG GAC GTG CAG CTG GTG ATG CTG GGC ACC GGC CCA CCT GAC CTG	1584
	513	G Q D V Q L V M L G T G P P D L	528
	1585	GAA CGA ATG CTG CAG CAC TTG GAG CGG GAG CAT CCC AAC AAG GTG CGC	1632
	529	E R M L Q H L E R E H P N K V R	544
	1633	GGG TGG GTC GGG TTC TCG GTC CTA ATG GTG CAT CGC ATC ACG CCG GGC	1680
10	545	G W V G F S V L M V H R I T P G	560
	1681	GCC AGC GTG CTG GTG ATG CCC TCC CGC TTC GCC GGC GGG CTG AAC CAG	1728
	561	A S V L V M P S R F A G G L N Q	576
	1729	CTC TAC GCG ATG GCA TAC GGC ACC GTC CCT GTG GTG CAC GCC GTG GGC	1776
	577	L Y A M A Y G T V P V V H A V G	592
15	1777	GGG CTC AGG GAC ACC GTG GCG CCG TTC GAC CCG TTC GGC GAC GCC GGG	1824
	593	G L R D T V A P F D P F G D A G	608
	1825	CTC GGG TGG ACT TTT GAC CGC GCC GAG GCC AAC AAG CTG ATC GAG GTG	1872
	609	L G W T F D R A E A N K L I E V	624
	1873	CTC AGC CAC TGC CTC GAC ACG TAC CGA AAC TAC GAG GAG AGC TGG AAG	1920
20	625	L S H C L D T Y R N Y E E S W K	640
	1921	AGT CTC CAG GCG CGC GGC ATG TCG CAG AAC CTC AGC TGG GAC CAC GCG	1968
	641	S L Q A R G M S Q N L S W D H A	656
	1969	GCT GAG CTC TAC GAG GAC GTC CTT GTC AAG TAC CAG TGG	2007
	657	A E L Y E D V L V K Y Q W	669

TABLE 3
DNA Sequence and Deduced Amino Acid Sequence of
The Soluble Starch Synthase IIb Gene in Maize
[SEQ ID NO:10 and SEQ ID NO: 11]

25 FILE NAME : MSS3FULL.DNA SEQUENCE : NORMAL 2097 BP
CODON TABLE : UNIV.TCN
SEQUENCE REGION : 1 - 2097
TRANSLATION REGION : 1 - 2097

*** DNA TRANSLATION ***

30	1	ATG CCG GGG GCA ATC TCT TCC TCG TCG TCG GCT TTT CTC CTC CCC GTC	48
	1	M P G A I S S S S S A F L L P V	16
	49	GCG TCC TCC TCG CCG CGG CGC AGG CGG GGC AGT GTG GGT GCT GCT CTG	96
	17	A S S S P R R R R G S V G A A L	32
35	97	CGC TCG TAC GGC TAC AGC GGC GCG GAG CTG CGG TTG CAT TGG GCG CGG	144
	33	R S Y G Y S G A E L R L H W A R	48
	145	CGG GGC CCG CCT CAG GAT GGA GCG GCG TCG GTA CGC GCC GCA GCG GCA	192
	49	R G P P Q D G A A S V R A A A A	64
	193	CCG GCC GGG GGC GAA AGC GAG GAG GCA GCG AAG AGC TCC TCC TCG TCC	240
	65	P A G G E S E E A A K S S S S S	80
40	241	CAG GCG GGC GCT GTT CAG GGC AGC ACG GCC AAG GCT GTG GAT TCT GCT	288

	81	Q	A	G	A	V	Q	G	S	T	A	K	A	V	D	S	A	96
	289	TCA	CCT	CCC	AAT	CCT	TTG	ACA	TCT	GCT	CCG	AAG	CAA	AGT	CAG	AGC	GCT	336
	97	S	P	P	N	P	L	T	S	A	P	K	Q	S	Q	S	A	112
5	337	GCA	ATG	CAA	AAC	GGA	ACG	AGT	GGG	GGC	AGC	AGC	GCG	AGC	ACC	GCC	GCG	384
	113	A	M	Q	N	G	T	S	G	G	S	S	A	S	T	A	A	128
	385	CCG	GTG	TCC	GGA	CCC	AAA	GCT	GAT	CAT	CCA	TCA	GCT	CCT	GTC	ACC	AAG	432
	129	P	V	S	G	P	K	A	D	H	P	S	A	P	V	T	K	144
	433	AGA	GAA	ATC	GAT	GCC	AGT	GCG	GTG	AAG	CCA	GAG	CCC	GCA	GGT	GAT	GAT	480
	145	R	E	I	D	A	S	A	V	K	P	E	P	A	G	D	D	160
10	481	GCT	AGA	CCG	GTG	GAA	AGC	ATA	GGC	ATC	GCT	GAA	CCG	GTG	GAT	GCT	AAG	528
	161	A	R	P	V	E	S	I	G	I	A	E	P	V	D	A	K	176
	529	GCT	GAT	GCA	GCT	CCG	GCT	ACA	GAT	GCG	GCG	GCG	AGT	GCT	CCT	TAT	GAC	576
	177	A	D	A	A	P	A	T	D	A	A	A	S	A	P	Y	D	192
15	577	AGG	GAG	GAT	AAT	GAA	CCT	GGC	CCT	TTG	GCT	GGG	CCT	AAT	GTG	ATG	AAC	624
	193	R	E	D	N	E	P	G	P	L	A	G	P	N	V	M	N	208
	625	GTC	GTC	GTG	GTG	GCT	TCT	GAA	TGT	GCT	CCT	TTC	TGC	AAG	ACA	GGT	GGC	672
	209	V	V	V	V	A	S	E	C	A	P	F	C	K	T	G	G	224
	673	CTT	GGA	GAT	GTC	GTG	GGT	GCT	TTG	CCT	AAG	GCT	CTG	GCG	AGG	AGA	GGA	720
	225	L	G	D	V	V	G	A	L	P	K	A	L	A	R	R	G	240
20	721	CAC	CGT	GTT	ATG	GTC	GTG	ATA	CCA	AGA	TAT	GGA	GAG	TAT	GCC	GAA	GCC	768
	241	H	R	V	M	V	V	I	P	R	Y	G	E	Y	A	E	A	256
	769	CGG	GAT	TTA	GGT	GTA	AGG	AGA	CGT	TAC	AAG	GTA	GCT	GGA	CAG	GAT	TCA	816
	257	R	D	L	G	V	R	R	R	Y	K	V	A	G	Q	D	S	272
25	817	GAA	GTT	ACT	TAT	TTT	CAC	TCT	TAC	ATT	GAT	GGA	GTT	GAT	TTT	GTA	TTC	864
	273	E	V	T	Y	F	H	S	Y	I	D	G	V	D	F	V	F	288
	865	GTA	GAA	GCC	CCT	CCC	TTC	CGG	CAC	CGG	CAC	AAT	AAT	ATT	TAT	GGG	GGA	912
	289	V	E	A	P	P	F	R	H	R	H	N	N	I	Y	G	G	304
	913	GAA	AGA	TTG	GAT	ATT	TTG	AAG	CGC	ATG	ATT	TTG	TTC	TGC	AAG	GCC	GCT	960
	305	E	R	L	D	I	L	K	R	M	I	L	F	C	K	A	A	320
30	961	GTT	GAG	GTT	CCA	TGG	TAT	GCT	CCA	TGT	GGC	GGT	ACT	GTC	TAT	GGT	GAT	1008
	321	V	E	V	P	W	Y	A	P	C	G	G	T	V	Y	G	D	336
	1009	GGC	AAC	TTA	GTT	TTC	ATT	GCT	AAT	GAT	TGG	CAT	ACC	GCA	CTT	CTG	CCT	1056
	337	G	N	L	V	F	I	A	N	D	W	H	T	A	L	L	P	352
35	1057	GTC	TAT	CTA	AAG	GCC	TAT	TAC	CGG	GAC	AAT	GGT	TTG	ATG	CAG	TAT	GCT	1104
	353	V	Y	L	K	A	Y	Y	R	D	N	G	L	M	Q	Y	A	368
	1105	CGC	TCT	GTG	CTT	GTG	ATA	CAC	AAC	ATT	GCT	CAT	CAG	GGT	CGT	GGC	CCT	1152
	369	R	S	V	L	V	I	H	N	I	A	H	Q	G	R	G	P	384
	1153	GTA	GAC	GAC	TTC	GTC	AAT	TTT	GAC	TTG	CCT	GAA	CAC	TAC	ATC	GAC	CAC	1200
	385	V	D	D	F	V	N	F	D	L	P	E	H	Y	I	D	H	400
40	1201	TTC	AAA	CTG	TAT	GAC	AAC	ATT	GGT	GGG	GAT	CAC	AGC	AAC	GTT	TTT	GCT	1248
	401	F	K	L	Y	D	N	I	G	G	D	H	S	N	V	F	A	416
	1249	GCG	GGG	CTG	AAG	ACG	GCA	GAC	CGG	GTG	GTG	ACC	GTT	AGC	AAT	GGC	TAC	1296
	417	A	G	L	K	T	A	D	R	V	V	T	V	S	N	G	Y	432
45	1297	ATG	TGG	GAG	CTG	AAG	ACT	TCG	GAA	GGC	GGG	TGG	GGC	CTC	CAC	GAC	ATC	1344
	433	M	W	E	L	K	T	S	E	G	G	W	G	L	H	D	I	448

	1345	ATA	AAC	CAG	AAC	GAC	TGG	AAG	CTG	CAG	GGC	ATC	GTG	AAC	GGC	ATC	GAC	1392
	449	I	N	Q	N	D	W	K	L	Q	G	I	V	N	G	I	D	464
	1393	ATG	AGC	GAG	TGG	AAC	CCC	GCT	GTG	GAC	GTG	CAC	CTC	CAC	TCC	GAC	GAC	1440
	465	M	S	E	W	N	P	A	V	D	V	H	L	H	S	D	D	480
5	1441	TAC	ACC	AAC	TAC	ACG	TTC	GAG	ACG	CTG	GAC	ACC	GGC	AAG	CGG	CAG	TGC	1488
	481	Y	T	N	Y	T	F	E	T	L	D	T	G	K	R	Q	C	496
	1489	AAG	GCC	GCC	CTG	CAG	CGG	CAG	CTG	GGC	CTG	CAG	GTC	CGC	GAC	GAC	GTG	1536
	497	K	A	A	L	Q	R	Q	L	G	L	Q	V	R	D	D	V	512
10	1537	CCA	CTG	ATC	GGG	TTC	ATC	GGG	CGG	CTG	GAC	CAC	CAG	AAG	GGC	GTG	GAC	1584
	513	P	L	I	G	F	I	G	R	L	D	H	Q	K	G	V	D	528
	1585	ATC	ATC	GCC	GAC	GCG	ATC	CAC	TGG	ATC	GCG	GGG	CAG	GAC	GTG	CAG	CTC	632
	529	I	I	A	D	A	I	H	W	I	A	G	Q	D	V	Q	L	544
	1633	GTG	ATG	CTG	GGC	ACC	GGG	CGG	GCC	GAC	CTG	GAG	GAC	ATG	CTG	CGG	CGG	1680
	545	V	M	L	G	T	G	R	A	D	L	E	D	M	L	R	R	560
15	1681	TTC	GAG	TCG	GAG	CAC	AGC	GAC	AAG	GTG	CGC	GCG	TGG	GTG	GGG	TTC	TCG	1728
	561	F	E	S	E	H	S	D	K	V	R	A	W	V	G	F	S	576
	1729	GTG	CCC	CTG	GCG	CAC	CGC	ATC	ACG	GCG	GGC	GCG	GAC	ATC	CTG	CTG	ATG	1776
	577	V	P	L	A	H	R	I	T	A	G	A	D	I	L	L	M	592
20	1777	CCG	TCG	CGG	TTC	GAG	CCG	TGC	GGG	CTG	AAC	CAG	CTC	TAC	GCC	ATG	GCG	1824
	593	P	S	R	F	E	P	C	G	L	N	Q	L	Y	A	M	A	608
	1825	TAC	GGG	ACC	GTG	CCC	GTG	GTG	CAC	GCC	GTG	GGG	GGG	CTC	CGG	GAC	ACG	1872
	609	Y	G	T	V	P	V	V	H	A	V	G	G	L	R	D	T	624
	1873	GTG	GCG	CCG	TTC	GAC	CCG	TTC	AAC	GAC	ACC	GGG	CTC	GGG	TGG	ACG	TTC	1920
	625	V	A	P	F	D	P	F	N	D	T	G	L	G	W	T	F	640
25	1921	GAC	CGC	GCG	GAG	GCG	AAC	CGG	ATG	ATC	GAC	GCG	CTC	TCG	CAC	TGC	CTC	1968
	641	D	R	A	E	A	N	R	M	I	D	A	L	S	H	C	L	656
	1969	ACC	ACG	TAC	CGG	AAC	TAC	AAG	GAG	AGC	TGG	CGC	GCC	TGC	AGG	GCG	CGC	2016
	657	T	T	Y	R	N	Y	K	E	S	W	R	A	C	R	A	R	672
30	2017	GGC	ATG	GCC	GAG	GAC	CTC	AGC	TGG	GAC	CAC	GCC	GCC	GTG	CTG	TAT	GAG	2064
	673	G	M	A	E	D	L	S	W	D	H	A	A	V	L	Y	E	688
	2065	GAC	GTG	CTC	GTC	AAG	GCG	AAG	TAC	CAG	TGG	TGA						2097
	689	D	V	L	V	K	A	K	Y	Q	W	*						699

TABLE 4
DNA and Deduced Amino Acid Sequence of
The Soluble Starch Synthase I Gene in Maize
[SEQ ID NO:12; SEQ ID NO: 13]

FILE NAME : MSS1FULL.DNA SEQUENCE : NORMAL 1752 BP

CODON TABLE : UNIV.TCN

SEQUENCE REGION : 1 - 1752

TRANSLATION REGION : 1 - 1752

	TGC GTC GCG GAG CTG AGC AGG GAG GGG CCC GCG CCG CGC CCG CTG CCA Cys Val Ala Glu Leu Ser Arg Glu Gly Pro Ala Pro Arg Pro Leu Pro 700 705 710 715	48
5	CCC GCG CTG CTG GCG CCC CCG CTC GTG CCC GGC TTC CTC GCG CCG CCG Pro Ala Leu Leu Ala Pro Pro Leu Val Pro Gly Phe Leu Ala Pro Pro 720 725 730	96
	GCC GAG CCC ACG GGT GAG CCG GCA TCG ACG CCG CCG CCC GTG CCC GAC Ala Glu Pro Thr Gly Glu Pro Ala Ser Thr Pro Pro Pro Val Pro Asp 735 740 745	144
10	GCC GGC CTG GGG GAC CTC GGT CTC GAA CCT GAA GGG ATT GCT GAA GGT Ala Gly Leu Gly Asp Leu Gly Leu Glu Pro Glu Gly Ile Ala Glu Gly 750 755 760	192
15	TCC ATC GAT AAC ACA GTA GTT GTG GCA AGT GAG CAA GAT TCT GAG ATT Ser Ile Asp Asn Thr Val Val Val Ala Ser Glu Gln Asp Ser Glu Ile 765 770 775	240
	GTG GTT GGA AAG GAG CAA GCT CGA GCT AAA GTA ACA CAA AGC ATT GTC Val Val Gly Lys Glu Gln Ala Arg Ala Lys Val Thr Gln Ser Ile Val 780 785 790 795	288
20	TTT GTA ACC GGC GAA GCT TCT CCT TAT GCA AAG TCT GGG GGT CTA GGA Phe Val Thr Gly Glu Ala Ser Pro Tyr Ala Lys Ser Gly Gly Leu Gly 800 805 810	336
	GAT GTT TGT GGT TCA TTG CCA GTT GCT CTT GCT GCT CGT GGT CAC CGT Asp Val Cys Gly Ser Leu Pro Val Ala Leu Ala Ala Arg Gly His Arg 815 820 825	384
25	GTG ATG GTT GTA ATG CCC AGA TAT TTA AAT GGT ACC TCC GAT AAG AAT Val Met Val Val Met Pro Arg Tyr Leu Asn Gly Thr Ser Asp Lys Asn 830 835 840	432
30	TAT GCA AAT GCA TTT TAC ACA GAA AAA CAC ATT CGG ATT CCA TGC TTT Tyr Ala Asn Ala Phe Tyr Thr Glu Lys His Ile Arg Ile Pro Cys Phe 845 850 855	480
	GGC GGT GAA CAT GAA GTT ACC TTC TTC CAT GAG TAT AGA GAT TCA GTT Gly Gly Glu His Glu Val Thr Phe Phe His Glu Tyr Arg Asp Ser Val 860 865 870 875	528
35	GAC TGG GTG TTT GTT GAT CAT CCC TCA TAT CAC AGA CCT GGA AAT TTA Asp Trp Val Phe Val Asp His Pro Ser Tyr His Arg Pro Gly Asn Leu 880 885 890	576
	TAT GGA GAT AAG TTT GGT GCT TTT GGT GAT AAT CAG TTC AGA TAC ACA Tyr Gly Asp Lys Phe Gly Ala Phe Gly Asp Asn Gln Phe Arg Tyr Thr 895 900 905	624
40	CTC CTT TGC TAT GCT GCA TGT GAG GCT CCT TTG ATC CTT GAA TTG GGA Leu Leu Cys Tyr Ala Ala Cys Glu Ala Pro Leu Ile Leu Glu Leu Gly 910 915 920	672
45	GGA TAT ATT TAT GGA CAG AAT TGC ATG TTT GTT GTC AAT GAT TGG CAT Gly Tyr Ile Tyr Gly Gln Asn Cys Met Phe Val Val Asn Asp Trp His 925 930 935	720
	GCC AGT CTA GTG CCA GTC CTT CTT GCT GCA AAA TAT AGA CCA TAT GGT Ala Ser Leu Val Pro Val Leu Leu Ala Ala Lys Tyr Arg Pro Tyr Gly 940 945 950 955	768
50	GTT TAT AAA GAC TCC CGC AGC ATT CTT GTA ATA CAT AAT TTA GCA CAT Val Tyr Lys Asp Ser Arg Ser Ile Leu Val Ile His Asn Leu Ala His 960 965 970	816

	CAG GGT GTA GAG CCT GCA AGC ACA TAT CCT GAC CTT GGG TTG CCA CCT	864
	Gln Gly Val Glu Pro Ala Ser Thr Tyr Pro Asp Leu Gly Leu Pro Pro	
	975 980 985	
5	GAA TGG TAT GGA GCT CTG GAG TGG GTA TTC CCT GAA TGG GCG AGG AGG	912
	Glu Trp Tyr Gly Ala Leu Glu Trp Val Phe Pro Glu Trp Ala Arg Arg	
	990 995 1000	
	CAT GCC CTT GAC AAG GGT GAG GCA GTT AAT TTT TTG AAA GGT GCA GTT	960
	His Ala Leu Asp Lys Gly Glu Ala Val Asn Phe Leu Lys Gly Ala Val	
	1005 1010 1015	
10	GTG ACA GCA GAT CGA ATC GTG ACT GTC AGT AAG GGT TAT TCG TGG GAG	1008
	Val Thr Ala Asp Arg Ile Val Thr Val Ser Lys Gly Tyr Ser Trp Glu	
	1020 1025 1030 1035	
15	GTC ACA ACT GCT GAA GGT GGA CAG GGC CTC AAT GAG CTC TTA AGC TCC	1056
	Val Thr Thr Ala Glu Gly Gly Gln Gly Leu Asn Glu Leu Leu Ser Ser	
	1040 1045 1050	
	AGA AAG AGT GTA TTA AAC GGA ATT GTA AAT GGA ATT GAC ATT AAT GAT	1104
	Arg Lys Ser Val Leu Asn Gly Ile Val Asn Gly Ile Asp Ile Asn Asp	
	1055 1060 1065	
20	TGG AAC CCT GCC ACA GAC AAA TGT ATC CCC TGT CAT TAT TCT GTT GAT	1152
	Trp Asn Pro Ala Thr Asp Lys Cys Ile Pro Cys His Tyr Ser Val Asp	
	1070 1075 1080	
	GAC CTC TCT GGA AAG GCC AAA TGT AAA GGT GCA TTG CAG AAG GAG CTG	1200
	Asp Leu Ser Gly Lys Ala Lys Cys Lys Gly Ala Leu Gln Lys Glu Leu	
	1085 1090 1095	
25	GGT TTA CCT ATA AGG CCT GAT GTT CCT CTG ATT GGC TTT ATT GGA AGG	1248
	Gly Leu Pro Ile Arg Pro Asp Val Pro Leu Ile Gly Phe Ile Gly Arg	
	1100 1105 1110 1115	
30	TTG GAT TAT CAG AAA GGC ATT GAT CTC ATT CAA CTT ATC ATA CCA GAT	1296
	Leu Asp Tyr Gln Lys Gly Ile Asp Leu Ile Gln Leu Ile Ile Pro Asp	
	1120 1125 1130	
	CTC ATG CGG GAA GAT GTT CAA TTT GTC ATG CTT GGA TCT GGT GAC CCA	1344
	Leu Met Arg Glu Asp Val Gln Phe Val Met Leu Gly Ser Gly Asp Pro	
	1135 1140 1145	
35	GAG CTT GAA GAT TGG ATG AGA TCT ACA GAG TCG ATC TTC AAG GAT AAA	1392
	Glu Leu Glu Asp Trp Met Arg Ser Thr Glu Ser Ile Phe Lys Asp Lys	
	1150 1155 1160	
	TTT CGT GGA TGG GTT GGA TTT AGT GTT CCA GTT TCC CAC CGA ATA ACT	1440
	Phe Arg Gly Trp Val Gly Phe Ser Val Pro Val Ser His Arg Ile Thr	
	1165 1170 1175	
40	GCC GGC TGC GAT ATA TTG TTA ATG CCA TCC AGA TTC GAA CCT TGT GGT	1488
	Ala Gly Cys Asp Ile Leu Leu Met Pro Ser Arg Phe Glu Pro Cys Gly	
	1180 1185 1190 1195	
45	CTC AAT CAG CTA TAT GCT ATG CAG TAT GGC ACA GTT CCT GTT GTC CAT	1536
	Leu Asn Gln Leu Tyr Ala Met Gln Tyr Gly Thr Val Pro Val Val His	
	1200 1205 1210	
	GCA ACT GGG GGC CTT AGA GAT ACC GTG GAG AAC TTC AAC CCT TTC GGT	1584
	Ala Thr Gly Gly Leu Arg Asp Thr Val Glu Asn Phe Asn Pro Phe Gly	
	1215 1220 1225	
50	GAG AAT GGA GAG CAG GGT ACA GGG TGG GCA TTC GCA CCC CTA ACC ACA	1632
	Glu Asn Gly Glu Gln Gly Thr Gly Trp Ala Phe Ala Pro Leu Thr Thr	
	1230 1235 1240	

GAA AAC ATG TTT GTG GAC ATT GCG AAC TGC AAT ATC TAC ATA CAG GGA 1680
 Glu Asn Met Phe Val Asp Ile Ala Asn Cys Asn Ile Tyr Ile Gln Gly
 1245 1250 1255

5 ACA CAA GTC CTC CTG GGA AGG GCT AAT GAA GCG AGG CAT GTC AAA AGA 1728
 Thr Gln Val Leu Leu Gly Arg Ala Asn Glu Ala Arg His Val Lys Arg
 1260 1265 1270 1275

CTT CAC GTG GGA CCA TGC CGC TGA 1752
 Leu His Val Gly Pro Cys Arg *
 1280

10 (2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 584 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Cys Val Ala Glu Leu Ser Arg Glu Gly Pro Ala Pro Arg Pro Leu Pro
 1 5 10 15
 20 Pro Ala Leu Leu Ala Pro Pro Leu Val Pro Gly Phe Leu Ala Pro Pro
 20 25 30
 Ala Glu Pro Thr Gly Glu Pro Ala Ser Thr Pro Pro Pro Val Pro Asp
 35 40 45
 Ala Gly Leu Gly Asp Leu Gly Leu Glu Pro Glu Gly Ile Ala Glu Gly
 50 55 60
 25 Ser Ile Asp Asn Thr Val Val Val Ala Ser Glu Gln Asp Ser Glu Ile
 65 70 75 80
 Val Val Gly Lys Glu Gln Ala Arg Ala Lys Val Thr Gln Ser Ile Val
 85 90 95
 30 Phe Val Thr Gly Glu Ala Ser Pro Tyr Ala Lys Ser Gly Gly Leu Gly
 100 105 110
 Asp Val Cys Gly Ser Leu Pro Val Ala Leu Ala Ala Arg Gly His Arg
 115 120 125
 Val Met Val Val Met Pro Arg Tyr Leu Asn Gly Thr Ser Asp Lys Asn
 130 135 140
 35 Tyr Ala Asn Ala Phe Tyr Thr Glu Lys His Ile Arg Ile Pro Cys Phe
 145 150 155 160
 Gly Gly Glu His Glu Val Thr Phe Phe His Glu Tyr Arg Asp Ser Val
 165 170 175
 40 Asp Trp Val Phe Val Asp His Pro Ser Tyr His Arg Pro Gly Asn Leu
 180 185 190
 Tyr Gly Asp Lys Phe Gly Ala Phe Gly Asp Asn Gln Phe Arg Tyr Thr
 195 200 205
 Leu Leu Cys Tyr Ala Ala Cys Glu Ala Pro Leu Ile Leu Glu Leu Gly
 210 215 220
 45 Gly Tyr Ile Tyr Gly Gln Asn Cys Met Phe Val Val Asn Asp Trp His
 225 230 235 240

	Ala	Ser	Leu	Val	Pro	Val	Leu	Leu	Ala	Ala	Lys	Tyr	Arg	Pro	Tyr	Gly	
					245					250					255		
	Val	Tyr	Lys	Asp	Ser	Arg	Ser	Ile	Leu	Val	Ile	His	Asn	Leu	Ala	His	
				260					265					270			
5	Gln	Gly	Val	Glu	Pro	Ala	Ser	Thr	Tyr	Pro	Asp	Leu	Gly	Leu	Pro	Pro	
			275					280					285				
	Glu	Trp	Tyr	Gly	Ala	Leu	Glu	Trp	Val	Phe	Pro	Glu	Trp	Ala	Arg	Arg	
		290					295					300					
10	His	Ala	Leu	Asp	Lys	Gly	Glu	Ala	Val	Asn	Phe	Leu	Lys	Gly	Ala	Val	
	305					310					315					320	
	Val	Thr	Ala	Asp	Arg	Ile	Val	Thr	Val	Ser	Lys	Gly	Tyr	Ser	Trp	Glu	
				325						330					335		
	Val	Thr	Thr	Ala	Glu	Gly	Gly	Gln	Gly	Leu	Asn	Glu	Leu	Leu	Ser	Ser	
				340					345					350			
15	Arg	Lys	Ser	Val	Leu	Asn	Gly	Ile	Val	Asn	Gly	Ile	Asp	Ile	Asn	Asp	
			355					360					365				
	Trp	Asn	Pro	Ala	Thr	Asp	Lys	Cys	Ile	Pro	Cys	His	Tyr	Ser	Val	Asp	
		370					375					380					
20	Asp	Leu	Ser	Gly	Lys	Ala	Lys	Cys	Lys	Gly	Ala	Leu	Gln	Lys	Glu	Leu	
	385					390					395					400	
	Gly	Leu	Pro	Ile	Arg	Pro	Asp	Val	Pro	Leu	Ile	Gly	Phe	Ile	Gly	Arg	
				405						410					415		
	Leu	Asp	Tyr	Gln	Lys	Gly	Ile	Asp	Leu	Ile	Gln	Leu	Ile	Ile	Pro	Asp	
				420					425					430			
25	Leu	Met	Arg	Glu	Asp	Val	Gln	Phe	Val	Met	Leu	Gly	Ser	Gly	Asp	Pro	
			435					440					445				
	Glu	Leu	Glu	Asp	Trp	Met	Arg	Ser	Thr	Glu	Ser	Ile	Phe	Lys	Asp	Lys	
		450					455					460					
30	Phe	Arg	Gly	Trp	Val	Gly	Phe	Ser	Val	Pro	Val	Ser	His	Arg	Ile	Thr	
	465					470					475					480	
	Ala	Gly	Cys	Asp	Ile	Leu	Leu	Met	Pro	Ser	Arg	Phe	Glu	Pro	Cys	Gly	
				485						490					495		
	Leu	Asn	Gln	Leu	Tyr	Ala	Met	Gln	Tyr	Gly	Thr	Val	Pro	Val	Val	His	
			500						505					510			
35	Ala	Thr	Gly	Gly	Leu	Arg	Asp	Thr	Val	Glu	Asn	Phe	Asn	Pro	Phe	Gly	
			515					520					525				
	Glu	Asn	Gly	Glu	Gln	Gly	Thr	Gly	Trp	Ala	Phe	Ala	Pro	Leu	Thr	Thr	
		530					535					540					
40	Glu	Asn	Met	Phe	Val	Asp	Ile	Ala	Asn	Cys	Asn	Ile	Tyr	Ile	Gln	Gly	
	545					550					555					560	
	Thr	Gln	Val	Leu	Leu	Gly	Arg	Ala	Asn	Glu	Ala	Arg	His	Val	Lys	Arg	
				565						570					575		
	Leu	His	Val	Gly	Pro	Cys	Arg	*									
			580														

TABLE 5
mRNA Sequence and Deduced Amino Acid Sequence of
The Maize Branching Enzyme II Gene and the Transit Peptide
[SEQ ID NO:14 and SEQ ID NO:15]

5	LOCUS	MZEGLUCTRN 2725 bp ss-mRNA	PLN
	DEFINITION	Corn starch branching enzyme II mRNA, complete cds.	
	ACCESSION	L08065	
	KEYWORDS	1,4-alpha-glucan branching enzyme; amylo-transglycosylase; glucanotransferase; starch branching enzyme II.	
10	SOURCE	Zea mays cDNA to mRNA.	
	ORGANISM	Zea mays Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida; Commelinidae; Cyperales; Poaceae.	
	REFERENCE	1 (bases 1 to 2725)	
15	AUTHORS	Fisher, D.K., Boyer, C.D. and Hannah, L.C.	
	TITLE	Starch branching enzyme II from maize endosperm	
	JOURNAL	Plant Physiol. 102, 1045-1046 (1993)	
	STANDARD	full automatic	
	COMMENT	NCBI gi: 168482	
20	FEATURES	Location/Qualifiers	
	source	1..2725 /cultivar="W64Ax182E" /dev_stage="29 days post pollination" /tissue_type="endosperm" /organism="Zea mays"	
25	sig_peptide	91..264 /codon_start=1	
	CDS	91..2490 /EC_number="2.4.1.18" /note="NCBI gi: 168483" /codon_start=1 /product="starch branching enzyme II"	
30		/translation="MAFRVSGAVLGGAVRAPRLTGGEGSLVFRHTGLFLTRGARVGC	
35		SGTHGAMRAAAAARKAVMVPEGEN DGLASRADSAQFQSDELEVDPDISEETTCGAGVAD	
		AQALNRVRVPPSPDGQKIFQIDPMLQGYKYHLEYRYSLYRRIRSDIDEHEGGLEAFS	
40		RSYEKFGFNASAE GITYREWAPGAFSAALVGDVNNWDPNADRM SKNEFGVWEIFLPNN	
		ADGTSPIPHGSRVKVRMDTPSGIKDSIPAWIKYSVQAPGEIPYDGIYYDPPEEVKYVF	
45		RHAQPKRPKSLRIYETHVGMSSPEPKINTYVNF RDEVLPRIKKLGYNAVQIMAIQEH S	
		YYGSFGYHVTNFFAPSSRFGTPEDL KSLIDRAHELGLLV LMDVVHSHASSNTLDGLNG	
		FDGTDTHYFHSGPRGHHMMWDSRLFN YGNWEVLRFLLSNARWWLEEYKFDGFRFDGVT	
50		SMMYTHHGLQVTF TGNFNEYFGFATDVDAVVYLM LVNDLIHGLYPEAVTIGEDVSGMP	
		TFALPVHDGGVGFDYRMHMAVADKWIDLLKQSD ETWKMGDIVHTLTNRRWLEKCVTYA	
55		ESHDQALVGDKTIAFWLMDKDMYDFMALDRPSTPTIDRGIALHKMIRLITMGLGEGY	
		LNFMGNEFGHPEWIDFPRGPQRLPSGKFIPGN NNSYDKCRRRFDLGADYLR YHGMQE	
		FDQAMQHLEQKYEFMTSDHQYISRKHEEDKVIVFEKGDLVFVFNHFCNNSYFDYRIGC	
60		RKPGVYKVVLDS DAGLFGGFSRIHHAAEHFTADCSHDNR PYSFSVYTPSRTC VVYAPV	
	mat_peptide	E" 265..2487 /codon_start=1 /product="starch branching enzyme II"	
65	BASE COUNT	727 A	534 C 715 G 749 T

ORIGIN

```

1  GGGCCAGAGC AGACCCGGAT TTCGCTCTTG CGGTCGCTGG GGTTTTAGCA TTGGCTGATC
61 AGTTTCGATCC GATCCGGCTG CGAAGGCGAG ATGGCGTTCC GGGTTTCTGG GCGCGTGCTC
121 GGTGGGGCCG TAAGGGCTCC CCGACTCACC GGCGGCGGGG AGGGTAGTCT AGTCTTCCGG
5  181 CACACCGGCC TCTTCTTAAC TCGGGGTGCT CGAGTTGGAT GTTCGGGGAC GCACGGGGCC
241 ATGCGCGCGG CGGCCGCGGC CAGGAAGGCG GTCATGGTTC CTGAGGGCGA GAATGATGGC
301 CTCGCATCAA GGGCTGACTC GGCTCAATTC CAGTCCGATG AACTGGAGGT ACCAGACATT
361 TCTGAAGAGA CAACGTGCGG TGCTGGTGTG GCTGATGCTC AAGCCTTGAA CAGAGTTCGA
10 421 GTGGTCCCCC CACCAAGCGA TGGACAAAAA ATATTCCAGA TTGACCCCAT GTTGCAAGGC
481 TATAAGTACC ATCTTGAGTA TCGGTACAGC CTCTATAGAA GAATCCGTTT AGACATTGAT
541 GAACATGAAG GAGGCTTGGA AGCCTTCTCC CGTAGTTATG AGAAGTTTGG ATTTAATGCC
601 AGCGCGGAAG GTATCACATA TCGAGAATGG GCTCCTGGAG CATTTTCTGC AGCATTGGTG
661 GGTGACGTCA ACAACTGGGA TCCAAATGCA GATCGTATGA GCAAAAATGA GTTTGGTGTG
721 TGGGAAATTT TTCTGCCTAA CAATGCAGAT GGTACATCAC CTATTCCTCA TGGATCTCGT
15 781 GTAAAGGTGA GAATGGATAC TCCATCAGGG TCCATCCAGC CTGGATCAAG
841 TACTCAGTGC AGGCCCCAGG AGAAATACCA TATGATGGGA TTTATTATGA TCCTCCTGAA
901 GAGGTAAAGT ATGTGTTTCA GCATGCGCAA CCTAAACGAC CAAAATCATT GCGGATATAT
961 GAAACACATG TCGGAATGAG TAGCCCGGAA CCGAAGATAA ACACATATGT AAACCTTAGG
1021 GATGAAGTCC TCCCAAGAAT AAAAAAATT GGATACAATG CAGTGCAAAAT AATGGCAATC
20 1081 CAAGAGCACT CATATTATGG AAGCTTTGGA TACCATGTAA CTAATTTTTT TGCGCCAAGT
1141 AGTCGTTTTG GTACCCAGAG AGATTTGAAG TCTTTGATTG ATAGAGCACA TGAGCTTGGT
1201 TTGCTAGTTC TCATGGATGT GGTTCATAGT CATGCGTCAA GTAATACTCT GGATGGGTTG
1261 AATGGTTTTG ATGGTACAGA TACACATTAC TTTCACAGTG GTCCACGTGG CCATCACTGG
1321 ATGTGGGATT CTCGCCTATT TAACATGTTG AACTGGGAAG TTTTAAGATT TCTTCTCTCC
25 1381 AATGCTAGAT GGTGGCTCGA GGAATATAAG TTTGATGGTT TCCGTTTTGA TGGTGTGACC
1441 TCCATGATGT ACACTACCA CGGATTACAA GTAACATTTA CGGGGAACCT CAATGAGTAT
1501 TTTGGCTTTG CCACCGATGT AGATGCAGTG GTTACTTGA TGCTGGTAAA TGATCTAATT
1561 CATGGACTTT ATCCTGAGGC TGTAACCATT GGTGAAGATG TTAGTGGAAT GCCTACATTT
1621 GCCCTTCTCTG TTCACGATGG TGGGGTAGGT TTTGACTATC GGATGCATAT GGCTGTGGCT
30 1681 GACAAATGGA TTGACCTTCT CAAGCAAAGT GATGAAACTT GGAAGATGGG TGATATTGTG
1741 CACACACTGA CAAATAGGAG GTGGTTAGAG AAGTGTGTAA CTTATGCTGA AAGTCATGAT
1801 CAAGCATTAG TCGGCGACAA GACTATTGCG TTTTGGTTGA TGGACAAGGA TATGTATGAT
1861 TTCATGGCCC TCGATAGACC TTCAACTCCT ACCATTGATC GTGGGATAGC ATTACATAAG
1921 ATGATTAGAC TTATCACAAT GGGTTTAGGA GGAGAGGGCT ATCTTAATTT CATGGGAAAT
35 1981 GAGTTTGGAC ATCCTGAATG GATAGATTTT CCAAGAGGTC CGCAAAGACT TCCAAGTGGT
2041 AAGTTTATTC CAGGGAATAA CAACAGTTAT GACAAATGTC GTCGAAGATT TGACCTGGGT
2101 GATGCAGACT ATCTTAGGTA TCATGGTATG CAAGAGTTTG ATCAGGCAAT GCAACATCTT
2161 GAGCAAAAT ATGAATTCAT GACATCTGAT CACCAGTATA TTTCCCGGAA ACATGAGGAG
2221 GATAAGGTGA TTGTGTTTGA AAAGGGAGAT TTGGTATTTG TGTTCAACTT CCACTGCAAC
40 2281 AACAGCTATT TTGACTACCG TATTGGTTGT CGAAAGCCTG GGGTGTATAA GGTGGTCTTG
2341 GACTCCGACG CTGGACTATT TGGTGGATTT AGCAGGATCC ATCAGCAGC CGAGCACTTC
2401 ACCGCCGACT GTTCGCATGA TAATGAGCCA TATTCATTCT CCGTTTATAC ACCAAGCAGA
2461 ACATGTGTCTG TCTATGCTCC AGTGGAGTGA TAGCGGGGTA CTCGTTGCTG CGCGGCATGT
45 2521 GTGGGGCTGT CGATGTGAGG AAAAACCTTC TTCCAAAACC GGCAGATGCA TGCATGCATG
2581 CTACAATAAG GTTCTGATAC TTTAATCGAT GCTGGAAAGC CCATGCATCT CGCTGCGTTG
2641 TCCTCTCTAT ATATATAAGA CCTTCAAGGT GTCAATTAAA CATAGAGTTT TCGTTTTTTCG
2701 CTTTCCTAAA AAAAAAAAAA AAAAA

```

//

TABLE 6

mRNA Sequence and Deduced Amino Acid Sequence of the
Maize Branching Enzyme I and the Transit Peptide
[SEQ ID NO:16 and SEQ ID NO:17]

50

55

60

```

LOCUS      MZEBEI      2763 bp ss-mRNA      PLN
DEFINITION Maize mRNA for branching enzyme-I (BE-I).
ACCESSION  D11081
KEYWORDS   branching enzyme-I.
SOURCE     Zea mays L. (inbred Oh43), cDNA to mRNA.
ORGANISM   Zea mays
            Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida;
            Commelinidae; Liliopsida.
REFERENCE  1 (bases 1 to 2763)
AUTHORS    Baba,T., Kimura,K., Mizuno,K., Etoh,H., Ishida,Y., Shida,O. and
            Arai,Y.

```

TITLE Sequence conservation of the catalytic regions of Amylolytic
 enzymes in maize branching enzyme-I
 JOURNAL Biochem. Biophys. Res. Commun. 181, 87-94 (1991)
 STANDARD full automatic
 5 COMMENT Submitted (30-APR-1992) to DDBJ by: Tadashi Bana
 Institute of Applied Biochemistry
 University of Tsukuba
 Tsukuba, Ibaraki 305
 Japan
 10 Phone: 0298-53-6632
 Fax: 0298-53-6632.

NCBI gi: 217959
 FEATURES Location/Qualifiers
 15 source 1..2763
 /organism="Zea mays"
 CDS <1..2470
 /note="NCBI gi: 217960"
 /codon_start=2
 /product="branching enzyme-I precursor"
 20 /translation="LCLVSPSSSPTPLPPRRSRSHADRAAPPGIAGGGNVRLSVLSV
 QCKARRSGVRKVSKFATAATVQEDKTMATAKGDVDHLPIDLDPKLEIFKDHFRYRM
 25 KRFLEQKGSIEENEGSLESFSKGYLKFGINTNEDGTVYREWAPAAQAEELIGDFNDWN
 GANHKMEKDKFGVWSIKIDHVKGKPAIPHNSKVKFRFLHGGVWVDRIPALIRYATVDA
 SKFGAPYDGVHWDPPASERYTFKHPRPSKPAAPRIYEAHVGMGGEKPAVSTYREFADN
 30 VLPRIRANNYNVQLMAVMEHSYYASFGYHVTNFFAVSSRSGTPEDLKYLVDKAHSLG
 LRVLMDEVHSHASNNVTDGLNGYDVGQSTQESYFHAGDRGYHKLWDSRLFNANWEVL
 35 RFLLSNLRYWLDEFMFDFRFDGVTSMLYHHHGINVGFTGNYQYFSLDTAVDAVVYM
 MLANHLMHKLLPEATVVAEDVSGMPVLCRPVDEGGVGFDYRLAMAIPDRWIDYLNKND
 DSEWSMGEIAHTLTNRRYTEKCIAYAESHDSIVGDKTIAFLMDKEMYTGMSDLQPA
 40 SPTIDRGIALQKMIHFITMALGGDGYLNFMGNEFGHPWIDFPREGNNWSYDKCRRQW
 SLVDTDHLRYKYMNAFDQAMNALDERFSFLSSSKQIVSDMNDEEKVIVFERGDLVVFV
 45 NFHPKKTYEGYKVGCDLPGKYRVALDSDALVFGGHGRVGHVDHFTSPEGVPGVPETN
 FNNRPNSFKVLSPPRTCVAIYRVDEAGAGRRHLHAKAETGKTSPAESIDVKASRASSKE
 DKEATAGGKKGWKFARQPSDQDTK"
 50 transit_peptide 2..190
 mat_peptide 191..2467
 /EC_number="2.4.1.18"
 /codon_start=1
 /product="branching enzyme-I precursor"
 polyA_signal 2734..2739
 55 BASE COUNT 719 A 585 C 737 G 722 T
 ORIGIN
 1 GCTGTGCCTC GTGTGCGCCT CTTCCTCGCC GACTCCGCTT CCGCCGCCGC GCGCTCTCG
 61 CTCGCATGCT GATCGGGCGG CACCGCCGGG GATCGGGGT GCGGCAATG TGCGCTGAG
 121 TGTGTTGTCT GTCCAGTGCA AGGCTCGCCG GTCAGGGGTG CGGAAGGTCA AGAGCAAATT
 60 181 CGCCACTGCA GCTACTGTGC AAGAAGATAA AACTATGGCA ACTGCCAAAG GCGATGTGCA
 241 CCATCTCCCC ATATACGACC TGGACCCCAA GCTGGAGATA TTCAAGGACC ATTCAGGTA
 301 CCGGATGAAA AGATTCTTAG AGCAGAAAGG ATCAATTGAA GAAAATGAGG GAAGTCTTGA
 361 ATCTTTTTCT AAAGGCTATT TGAATTGTTG GATTAATACA AATGAGGATG GAACTGTATA
 421 TCGTGAATGG GCACCTGCTG CGCAGGAGGC AGAGCTTATT GGTGACTTCA ATGACTGGAA
 65 481 TGGTGCAAAC CATAAGATGG AGAAGGATAA ATTTGGTGTT TGGTCGATCA AAATTGACCA
 541 TGTCAAAGGG AAACCTGCCA TCCCTCACAA TTCCAAGGTT AAATTTGCTT TTCTACATGG
 601 TGGAGTATGG GTTGATCGTA TTCCAGCATT GATTCGTTAT GCGACTGTTG ATGCCTCTAA

5 661 ATTTGGAGCT CCCTATGATG GTGTTTCATTG GGATCCTCCT GCTTCTGAAA GGTACACATT
721 TAAGCATCCT CGGCCTTCAA AGCCTGCTGC TCCACGTATC TATGAAGCCC ATGTAGGTAT
781 GAGTGGTGAA AAGCCAGCAG TAAGCACATA TAGGGAATTT GCAGACAATG TGTGGCCACG
841 CATACGAGCA AATAACTACA ACACAGTTCA GTTGATGGCA GTTATGGAGC ATTCGTAATA
901 TGCTTCTTTC GGGTACCATG TGACAAATTT CTTTGCAGTT AGCAGCAGAT CAGGCACACC
961 AGAGGACCTC AAATATCTTG TTGATAAGGC ACACAGTTTG GGTTCGCGAG TTCTGATGGA
1021 TGTGTGCCAT AGCCATGCAA GTAATAATGT CACAGATGGT TTAATGGCT ATGATGTTGG
1081 ACAAAGCACC CAAGAGTCCT ATTTTCATGC GGGAGATAGA GGTTATCATA AACTTTGGGA
1141 TAGTCGGCTG TTCAACTATG CTAACCTGGG GGTATTAAGG TTTCTTCTTT CTAACCTGAG
1201 ATATTGGTTG GATGAATTCA TGTTCGATGG CTTCGATTT GATGGAGTTA CATCAATGCT
1261 GTATCATCAC CATGGTATCA ATGTGGGGTT TACTGGAAAC TACCAAGAAAT ATTTTCAGTTT
1321 GGACACAGCT GTGGATGCAG TTGTTTACAT GATGCTTGCA AACCATTAA TGCACAAACT
1381 CTTGCCAGAA GCAACTGTTG TTGCTGAAGA TGTTTCAGGC ATGCCGGTCC TTTGCCGGCC
1441 AGTTGATGAA GGTGGGGTTG GGTTCGACTA TCGCCTGGCA ATGGCTATCC CTGATAGATG
1501 GATTGACTAC CTGAAGAATA AAGATGACTC TGAGTGGTCG ATGGGTGAAA TAGCGCATAC
1561 TTTGACTAAC AGGAGATATA CTGAAAAATG CATCGCATAT GCTGAGAGCC ATGATCAGTC
1621 TATTGTTGGC GACAAAACTA TTGCATTTCT CCTGATGGAC AAGGAAATGT AACTGGCAT
1681 GTCAGACTTG CAGCCTGCTT CACCTACAAT TGATCGAGGG ATTGCACTCC AAAAGATGAT
1741 TCACTTCATC ACAATGGCCC TTGGAGGTGA TGGCTACTTG AATTTTATGG GAAATGAGTT
1801 TGGTCACCCA GAATGGATTG ACTTTCCAAG AGAAGGGAAC AACTGGAGCT ATGATAAATG
1861 CAGACGACAG TGGAGCCTTG TGGACACTGA TCACTTGCAG TACAAGTACA TGAATGCGTT
1921 TGACCAAGCG ATGAATGCGC TCGATGAGAG ATTTTCCTTC CTTTCGTCGT CAAAGCAGAT
1981 CGTCAGCGAC ATGAACGATG AGGAAAAGGT TATTGTCTTT GAACGTGGAG ATTTAGTTTT
2041 TGTTTTCAAT TTCCATCCCA AGAAAACTTA CGAGGGCTAC AAAGTGGGAT GCGATTTGCC
2101 TGGGAAATAC AGAGTAGCCC TGGACTCTGA TGCTCTGGTC TTCGGTGGAC ATGGAAGAGT
2161 TGGCCACGAC GTGGATCACT TCACGTCGCC TGAAGGGGTG CCAGGGGTGC CCGAAACGAA
2221 CTTCAACAAC CGGCCGAACG CGTTCAAAGT CCTTCTCCG CCCCGCACCT GTGTGGCTTA
2281 TTACCGTGTA GACGAAGCAG GGGCTGGACG ACGTCTTAC GCGAAAGCAG AGACAGGAAA
2341 GACGTCTCCA GCAGAGAGCA TCGACGTCAA AGCTTCCAGA GCTAGTAGCA AAGAAGACAA
2401 GGAGGCAACG GCTGGTGGCA AGAAGGGATG GAAGTTTGC CCGCAGCCAT CCGATCAAGA
2461 TACCAATGA AGCCACGAGT CCTTGGTGAG GACTGGACTG GCTGCCGGCG CCCTGTTAGT
2521 AGTCCTGCTC TACTGGACTA GCCCGCGCTG GCGCCCTTGG AACGGTCCTT TCCTGTAGCT
2581 TGCAGGCGAC TGGTGTCTCA TCACCGAGCA GGCAGGCACT GCTTGTATAG CTTTTCTAGA
2641 ATAATAATCA GGGATGGATG GATGGTGTGT ATTGGCTATC TGGCTAGACG TGCATGTGCC
2701 CAGTTTGTAT GTACAGGAGC AGTTCCCGTC CAGAATAAAA AAAAATTGTG TGGGGGGTTT
2761 TTC

//

TABLE 7
Coding Sequence and Deduced Amino Acid Sequence for
Transit Peptide Region of the
Soluble Starch Synthase I Maize Gene (153 bp)
[SEQ ID NO:18 and SEQ ID NO:19]

40 FILE NAME : MSS1TRPT.DNA SEQUENCE : NORMAL 153 BP
CODON TABLE : UNIV.TCN
45 SEQUENCE REGION : 1 - 153
TRANSLATION REGION : 1 - 153
*** DNA TRANSLATION ***
1 ATG GCG ACG CCC TCG GCC GTG GGC GCC GCG TGC CTC CTC CTC GCG CGG 48
1 M A T P S A V G A A C L L L A R 16
50 49 GCC GCC TGG CCG GCC GCC GTC GGC GAC CGG GCG CGC CCG CGG AGG CTC 96
17 A A W P A A V G D R A R P R R L 32
97 CAG CGC GTG CTG CGC CGC CGG TGC GTC GCG GAG CTG AGC AGG GAG GGG 144
33 Q R V L R R R C V A E L S R E G 48
145 CCC CAT ATG 153
55 49 P H M 51

GFP constructs:

1. GFP only in pET-21a:

pEXS115 is digested with *Nde* I and *Xho* I and the 740 bp fragment containing the SGFP coding sequence is subcloned into the *Nde* I and *Xho* I sites of pET-21a (Novagen 601 Science Dr. Madison WI). (See FIG. 2b GFP-21a map.)

2. GFP subcloned in-frame at the 5' end of full-length mature WX:

The 740 bp *Nde* I fragment containing SGFP from pEXS114 is subcloned into the *Nde* I site of pEXSWX. (See FIG.3a GFP-FLWX map.)

3. GFP subcloned in-frame at the 5' end of N-terminally truncated WX:

WX truncated by 700 bp at N-terminus.

The 1 kb *Bam*H I fragment encoding the C-terminus of WX from pEXSWX is subcloned into the *Bgl* II site of pEXS115. Then the entire SGFP-truncated WX fragment is subcloned into pET21a as a *Nde* I-*Hind*III fragment. (See FIG. 3b GFP-BamHIWX map.)

4. GFP subcloned in-frame at the 5' end of truncated WX: WX truncated by 100 bp at N-terminus.

The 740 bp *Nde* I-*Nco* I fragment containing SGFP from pEXS115 is subcloned into pEXSWX at the *Nde* I and *Nco* I sites. (See Fig. 4 GFP-NcoWX map.)

Example Three:

Plasmid Transformation into Bacteria:

Escherichia coli competent cell preparation:

1. Inoculate 2.5 ml LB media with a single colony of desired *E. coli* strain : selected strain was XLIBLUE DL2IDE3 from (Stratagene); included appropriate antibiotics. Grow at 37°C, 250 rpm overnight.

2. Inoculate 100 ml of LB media with a 1:50 dilution of the overnight culture, including appropriate antibiotics. Grow at 37°C, 250 rpm until OD₆₀₀=0.3-0.5.

3. Transfer culture to sterile centrifuge bottle and chill on ice for 15 minutes.

4. Centrifuge 5 minutes at 3,000x g (4°C).
5. Resuspend pellet in 8 ml ice-cold Transformation buffer. Incubate on ice for 15 minutes.
6. Centrifuge 5 minutes at 3,000x g (4°C).
- 5 7. Resuspend pellet in 8 ml ice-cold Transformation buffer 2. Aliquot, flash-freeze in liquid nitrogen, and stored at -70°C.

	<u>Transformation Buffer 1</u>	<u>Transformation Buffer 2</u>
	RbCl 1.2 g	MOPS (10 mM) 0.209 g
	MnCl ₂ 4H ₂ O 0.99g	RbCl 0.12 g
10	K-Acetate 0.294 g	CaCl ₂ 2H ₂ O 1.1 g
	CaCl ₂ 2H ₂ O 0.15 g	Glycerol 15 g
	Glycerol 15 g	dH ₂ O 100 ml
	dH ₂ O 100 ml	pH to 6.8 with NaOH
	pH to 5.8 with 0.2 M acetic acid	Filter sterilize
15	Filter sterilize	

Escherichia coli transformation by rubidium chloride heat shock method: Hanahan, D. (1985) in DNA cloning: a practical approach (Glover, D.M. ed.), pp. 109-135, IRL Press.

1. Incubate 1-5 µl of DNA on ice with 150 µl *E. coli* competent cells for 30 minutes.
- 20 2. Heat shock at 42°C for 45 seconds.
3. Immediately place on ice for 2 minutes.
4. Add 600 µl LB media and incubate at 37°C for 1 hour.

5. Plate on LB agar including the appropriate antibiotics.

This plasmid will express the hybrid polypeptide containing the green fluorescent protein within the bacteria.

Example Four:

5 **Expression of Construct in *E. coli*:**

1. Inoculate 3 ml LB with *E. coli* containing plasmid of interest. Include appropriate antibiotics. 37°C, 250 rpm, overnight.
2. Inoculate 100 ml LB with 2 ml of overnight culture. Include appropriate antibiotics. Grow at 37°C, 250 rpm.
- 10 3. At OD₆₀₀ about 0.4-0.5, place at room temperature, 200 rpm.
4. At OD₆₀₀ about 0.6-0.8, induce with 100 µl 1M IPTG. Final IPTG concentration is 1 mM.
5. Grow at room temperature, 200 rpm, 4-5 hours.
6. Collect cells by centrifugation.
- 15 7. Flash freeze in liquid nitrogen and store at -70°C until use.

Cells can be resuspended in dH₂O and viewed under UV light ($\lambda_{\text{max}} = 395 \text{ nm}$) for intrinsic fluorescence. Alternatively, the cells can be sonicated and an aliquot of the cell extract can be separated by SDS-PAGE and viewed under UV light to detect GFP fluorescence. When the protein employed is a green fluorescent protein, the presence of the protein in the lysed material can be evaluated under UV at 395 nm in a light box and the signature green glow can be identified.

20

Example Five:

Plasmid Extraction from Bacteria:

The following is one of many common alkaline lysis plasmid purification protocols useful in practicing this invention.

- 5 1. Inoculate 100-200 ml LB media with a single colony of *E. coli* transformed with the one of the plasmids described above. Include appropriate antibiotics. Grow at 37°C, 250 rpm overnight.
2. Centrifuge 10 minutes at 5,000x g (4°C).
3. Resuspend cells in 10 ml water, transfer to a 15 ml centrifuge tube, and repeat
10 centrifugation.
4. Resuspend pellet in 5 ml 0.1 M NaOH, 0.5% SDS. Incubate on ice for 10 minutes.
5. Add 2.5 ml of 3 M sodium acetate (pH 5.2), invert gently, and incubate 10 minutes on ice.
6. Centrifuge 5 minutes at 15,000-20,000x g (4°C).
- 15 7. Extract supernatant with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1).
8. Centrifuge 10 minutes at 6,000-10,000x g (4°C).
9. Transfer aqueous phase to clean tube and precipitate with 1 volume of isopropanol.
10. Centrifuge 15 minutes at 12,000x g (4°C).
- 20 11. Dissolve pellet in 0.5 ml TE, add 20 µl of 10 mg/ml Rnase, and incubate 1 hour at 37°C.

12. Extract twice with phenol:chloroform:isoamyl alcohol (25:24:1).
13. Extract once with chloroform.
14. Precipitate aqueous phase with 1 volume of isopropanol and 0.1 volume of 3 M sodium acetate.
- 5 15. Wash pellet once with 70% ethanol.
16. Dry pellet in SpeedVac and resuspend pellet in TE.

This plasmid can then be inserted into other hosts.

TABLE 8
DNA Sequence and Deduced Amino Acid Sequence of
10 Starch Synthase Coding Region from pEXS52 [SEQ ID NO:20; SEQ ID NO:21]

FILE NAME : MSS1DELN.DNA SEQUENCE : NORMAL 1626 BP

CODON TABLE : UNIV.TCN

SEQUENCE REGION : 1 - 1626

TRANSLATION REGION : 1 - 1626

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

	TGC GTC GCG GAG CTG AGC AGG GAG GAC CTC GGT CTC GAA CCT GAA GGG	48
	Cys Val Ala Glu Leu Ser Arg Glu Asp Leu Gly Leu Glu Pro Glu Gly	
	55 60 65	
	ATT GCT GAA GGT TCC ATC GAT AAC ACA GTA GTT GTG GCA AGT GAG CAA	96
20	Ile Ala Glu Gly Ser Ile Asp Asn Thr Val Val Val Ala Ser Glu Gln	
	70 75 80	
	GAT TCT GAG ATT GTG GTT GGA AAG GAG CAA GCT CGA GCT AAA GTA ACA	144
	Asp Ser Glu Ile Val Val Gly Lys Glu Gln Ala Arg Ala Lys Val Thr	
	85 90 95	
25	CAA AGC ATT GTC TTT GTA ACC GGC GAA GCT TCT CCT TAT GCA AAG TCT	192
	Gln Ser Ile Val Phe Val Thr Gly Glu Ala Ser Pro Tyr Ala Lys Ser	
	100 105 110 115	
	GGG GGT CTA GGA GAT GTT TGT GGT TCA TTG CCA GTT GCT CTT GCT GCT	240
	Gly Gly Leu Gly Asp Val Cys Gly Ser Leu Pro Val Ala Leu Ala Ala	
30	120 125 130	
	CGT GGT CAC CGT GTG ATG GTT GTA ATG CCC AGA TAT TTA AAT GGT ACC	288
	Arg Gly His Arg Val Met Val Val Met Pro Arg Tyr Leu Asn Gly Thr	

	135	140	145	
	TCC GAT AAG AAT TAT GCA AAT GCA TTT TAC ACA GAA AAA CAC ATT CGG Ser Asp Lys Asn Tyr Ala Asn Ala Phe Tyr Thr Glu Lys His Ile Arg 150 155 160	336		
5	ATT CCA TGC TTT GGC GGT GAA CAT GAA GTT ACC TTC TTC CAT GAG TAT Ile Pro Cys Phe Gly Gly Glu His Glu Val Thr Phe Phe His Glu Tyr 165 170 175	384		
10	AGA GAT TCA GTT GAC TGG GTG TTT GTT GAT CAT CCC TCA TAT CAC AGA Arg Asp Ser Val Asp Trp Val Phe Val Asp His Pro Ser Tyr His Arg 180 185 190	432		
	CCT GGA AAT TTA TAT GGA GAT AAG TTT GGT GCT TTT GGT GAT AAT CAG Pro Gly Asn Leu Tyr Gly Asp Lys Phe Gly Ala Phe Gly Asp Asn Gln 200 205 210	480		
15	TTC AGA TAC ACA CTC CTT TGC TAT GCT GCA TGT GAG GCT CCT TTG ATC Phe Arg Tyr Thr Leu Leu Cys Tyr Ala Ala Cys Glu Ala Pro Leu Ile 215 220 225	528		
	CTT GAA TTG GGA GGA TAT ATT TAT GGA CAG AAT TGC ATG TTT GTT GTC Leu Glu Leu Gly Gly Tyr Ile Tyr Gly Gln Asn Cys Met Phe Val Val 230 235 240	576		
20	AAT GAT TGG CAT GCC AGT CTA GTG CCA GTC CTT CTT GCT GCA AAA TAT Asn Asp Trp His Ala Ser Leu Val Pro Val Leu Leu Ala Ala Lys Tyr 245 250 255	624		
25	AGA CCA TAT GGT GTT TAT AAA GAC TCC CGC AGC ATT CTT GTA ATA CAT Arg Pro Tyr Gly Val Tyr Lys Asp Ser Arg Ser Ile Leu Val Ile His 260 265 270 275	672		
	AAT TTA GCA CAT CAG GGT GTA GAG CCT GCA AGC ACA TAT CCT GAC CTT Asn Leu Ala His Gln Gly Val Glu Pro Ala Ser Thr Tyr Pro Asp Leu 280 285 290	720		
30	GGG TTG CCA CCT GAA TGG TAT GGA GCT CTG GAG TGG GTA TTC CCT GAA Gly Leu Pro Pro Glu Trp Tyr Gly Ala Leu Glu Trp Val Phe Pro Glu 295 300 305	768		
	TGG GCG AGG AGG CAT GCC CTT GAC AAG GGT GAG GCA GTT AAT TTT TTG Trp Ala Arg Arg His Ala Leu Asp Lys Gly Glu Ala Val Asn Phe Leu 310 315 320	816		
35	AAA GGT GCA GTT GTG ACA GCA GAT CGA ATC GTG ACT GTC AGT AAG GGT Lys Gly Ala Val Val Thr Ala Asp Arg Ile Val Thr Val Ser Lys Gly 325 330 335	864		
40	TAT TCG TGG GAG GTC ACA ACT GCT GAA GGT GGA CAG GGC CTC AAT GAG Tyr Ser Trp Glu Val Thr Thr Ala Glu Gly Gln Gly Leu Asn Glu 340 345 350 355	912		
	CTC TTA AGC TCC AGA AAG AGT GTA TTA AAC GGA ATT GTA AAT GGA ATT Leu Leu Ser Ser Arg Lys Ser Val Leu Asn Gly Ile Val Asn Gly Ile 360 365 370	960		
45	GAC ATT AAT GAT TGG AAC CCT GCC ACA GAC AAA TGT ATC CCC TGT CAT Asp Ile Asn Asp Trp Asn Pro Ala Thr Asp Lys Cys Ile Pro Cys His 375 380 385	1008		
	TAT TCT GTT GAT GAC CTC TCT GGA AAG GCC AAA TGT AAA GGT GCA TTG Tyr Ser Val Asp Asp Leu Ser Gly Lys Ala Lys Cys Lys Gly Ala Leu 390 395 400	1056		

	CAG AAG GAG CTG GGT TTA CCT ATA AGG CCT GAT GTT CCT CTG ATT GGC	1104
	Gln Lys Glu Leu Gly Leu Pro Ile Arg Pro Asp Val Pro Leu Ile Gly	
	405 410 415	
5	TTT ATT GGA AGG TTG GAT TAT CAG AAA GGC ATT GAT CTC ATT CAA CTT	1152
	Phe Ile Gly Arg Leu Asp Tyr Gln Lys Gly Ile Asp Leu Ile Gln Leu	
	420 425 430 435	
	ATC ATA CCA GAT CTC ATG CGG GAA GAT GTT CAA TTT GTC ATG CTT GGA	1200
	Ile Ile Pro Asp Leu Met Arg Glu Asp Val Gln Phe Val Met Leu Gly	
	440 445 450	
10	TCT GGT GAC CCA GAG CTT GAA GAT TGG ATG AGA TCT ACA GAG TCG ATC	1248
	Ser Gly Asp Pro Glu Leu Glu Asp Trp Met Arg Ser Thr Glu Ser Ile	
	455 460 465	
15	TTC AAG GAT AAA TTT CGT GGA TGG GTT GGA TTT AGT GTT CCA GTT TCC	1296
	Phe Lys Asp Lys Phe Arg Gly Trp Val Gly Phe Ser Val Pro Val Ser	
	470 475 480	
	CAC CGA ATA ACT GCC GGC TGC GAT ATA TTG TTA ATG CCA TCC AGA TTC	1344
	His Arg Ile Thr Ala Gly Cys Asp Ile Leu Leu Met Pro Ser Arg Phe	
	485 490 495	
20	GAA CCT TGT GGT CTC AAT CAG CTA TAT GCT ATG CAG TAT GGC ACA GTT	1392
	Glu Pro Cys Gly Leu Asn Gln Leu Tyr Ala Met Gln Tyr Gly Thr Val	
	500 505 510 515	
	CCT GTT GTC CAT GCA ACT GGG GGC CTT AGA GAT ACC GTG GAG AAC TTC	1440
	Pro Val Val His Ala Thr Gly Gly Leu Arg Asp Thr Val Glu Asn Phe	
	520 525 530	
25	AAC CCT TTC GGT GAG AAT GGA GAG CAG GGT ACA GGG TGG GCA TTC GCA	1488
	Asn Pro Phe Gly Glu Asn Gly Glu Gln Gly Thr Gly Trp Ala Phe Ala	
	535 540 545	
30	CCC CTA ACC ACA GAA AAC ATG TTT GTG GAC ATT GCG AAC TGC AAT ATC	1536
	Pro Leu Thr Thr Glu Asn Met Phe Val Asp Ile Ala Asn Cys Asn Ile	
	550 555 560	
	TAC ATA CAG GGA ACA CAA GTC CTC CTG GGA AGG GCT AAT GAA GCG AGG	1584
	Tyr Ile Gln Gly Thr Gln Val Leu Leu Gly Arg Ala Asn Glu Ala Arg	
	565 570 575	
35	CAT GTC AAA AGA CTT CAC GTG GGA CCA TGC CGC TGA	1620
	His Val Lys Arg Leu His Val Gly Pro Cys Arg *	
	580 585 590	

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

40 (A) LENGTH: 540 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

45 Cys Val Ala Glu Leu Ser Arg Glu Asp Leu Gly Leu Glu Pro Glu Gly
1 5 10 15
Ile Ala Glu Gly Ser Ile Asp Asn Thr Val Val Val Ala Ser Glu Gln
20 25 30
Asp Ser Glu Ile Val Val Gly Lys Glu Gln Ala Arg Ala Lys Val Thr
35 40 45

	Gln	Ser	Ile	Val	Phe	Val	Thr	Gly	Glu	Ala	Ser	Pro	Tyr	Ala	Lys	Ser	
	50						55					60					
	Gly	Gly	Leu	Gly	Asp	Val	Cys	Gly	Ser	Leu	Pro	Val	Ala	Leu	Ala	Ala	
	65					70					75					80	
5	Arg	Gly	His	Arg	Val	Met	Val	Val	Met	Pro	Arg	Tyr	Leu	Asn	Gly	Thr	
					85					90					95		
	Ser	Asp	Lys	Asn	Tyr	Ala	Asn	Ala	Phe	Tyr	Thr	Glu	Lys	His	Ile	Arg	
				100					105					110			
10	Ile	Pro	Cys	Phe	Gly	Gly	Glu	His	Glu	Val	Thr	Phe	Phe	His	Glu	Tyr	
			115					120					125				
	Arg	Asp	Ser	Val	Asp	Trp	Val	Phe	Val	Asp	His	Pro	Ser	Tyr	His	Arg	
		130					135					140					
	Pro	Gly	Asn	Leu	Tyr	Gly	Asp	Lys	Phe	Gly	Ala	Phe	Gly	Asp	Asn	Gln	
	145					150					155					160	
15	Phe	Arg	Tyr	Thr	Leu	Leu	Cys	Tyr	Ala	Ala	Cys	Glu	Ala	Pro	Leu	Ile	
					165					170					175		
	Leu	Glu	Leu	Gly	Gly	Tyr	Ile	Tyr	Gly	Gln	Asn	Cys	Met	Phe	Val	Val	
				180					185					190			
20	Asn	Asp	Trp	His	Ala	Ser	Leu	Val	Pro	Val	Leu	Leu	Ala	Ala	Lys	Tyr	
			195					200					205				
	Arg	Pro	Tyr	Gly	Val	Tyr	Lys	Asp	Ser	Arg	Ser	Ile	Leu	Val	Ile	His	
		210					215					220					
	Asn	Leu	Ala	His	Gln	Gly	Val	Glu	Pro	Ala	Ser	Thr	Tyr	Pro	Asp	Leu	
	225					230					235					240	
25	Gly	Leu	Pro	Pro	Glu	Trp	Tyr	Gly	Ala	Leu	Glu	Trp	Val	Phe	Pro	Glu	
					245					250					255		
	Trp	Ala	Arg	Arg	His	Ala	Leu	Asp	Lys	Gly	Glu	Ala	Val	Asn	Phe	Leu	
				260					265					270			
30	Lys	Gly	Ala	Val	Val	Thr	Ala	Asp	Arg	Ile	Val	Thr	Val	Ser	Lys	Gly	
		275						280					285				
	Tyr	Ser	Trp	Glu	Val	Thr	Thr	Ala	Glu	Gly	Gly	Gln	Gly	Leu	Asn	Glu	
		290					295					300					
	Leu	Leu	Ser	Ser	Arg	Lys	Ser	Val	Leu	Asn	Gly	Ile	Val	Asn	Gly	Ile	
	305					310					315					320	
35	Asp	Ile	Asn	Asp	Trp	Asn	Pro	Ala	Thr	Asp	Lys	Cys	Ile	Pro	Cys	His	
					325					330					335		
	Tyr	Ser	Val	Asp	Asp	Leu	Ser	Gly	Lys	Ala	Lys	Cys	Lys	Gly	Ala	Leu	
				340					345					350			
40	Gln	Lys	Glu	Leu	Gly	Leu	Pro	Ile	Arg	Pro	Asp	Val	Pro	Leu	Ile	Gly	
		355						360					365				
	Phe	Ile	Gly	Arg	Leu	Asp	Tyr	Gln	Lys	Gly	Ile	Asp	Leu	Ile	Gln	Leu	
		370					375					380					
	Ile	Ile	Pro	Asp	Leu	Met	Arg	Glu	Asp	Val	Gln	Phe	Val	Met	Leu	Gly	
	385					390					395					400	
45	Ser	Gly	Asp	Pro	Glu	Leu	Glu	Asp	Trp	Met	Arg	Ser	Thr	Glu	Ser	Ile	
					405					410					415		

	Phe	Lys	Asp	Lys	Phe	Arg	Gly	Trp	Val	Gly	Phe	Ser	Val	Pro	Val	Ser
				420					425					430		
	His	Arg	Ile	Thr	Ala	Gly	Cys	Asp	Ile	Leu	Leu	Met	Pro	Ser	Arg	Phe
			435					440					445			
5	Glu	Pro	Cys	Gly	Leu	Asn	Gln	Leu	Tyr	Ala	Met	Gln	Tyr	Gly	Thr	Val
		450					455					460				
	Pro	Val	Val	His	Ala	Thr	Gly	Gly	Leu	Arg	Asp	Thr	Val	Glu	Asn	Phe
	465					470					475					480
10	Asn	Pro	Phe	Gly	Glu	Asn	Gly	Glu	Gln	Gly	Thr	Gly	Trp	Ala	Phe	Ala
				485						490					495	
	Pro	Leu	Thr	Thr	Glu	Asn	Met	Phe	Val	Asp	Ile	Ala	Asn	Cys	Asn	Ile
				500					505					510		
	Tyr	Ile	Gln	Gly	Thr	Gln	Val	Leu	Leu	Gly	Arg	Ala	Asn	Glu	Ala	Arg
			515					520					525			
15	His	Val	Lys	Arg	Leu	His	Val	Gly	Pro	Cys	Arg	*				
		530					535					540				

Example Six:

This experiment employs a plasmid having a maize promoter, a maize transit peptide, a starch-encapsulating region from the starch synthase I gene, and a ligated gene fragment attached thereto. The plasmid shown in FIG. 6 contains the DNA sequence listed in Table 8.

Plasmid pEXS52 was constructed according to the following protocol:

Materials used to construct transgenic plasmids are as follows:

- Plasmid pBluescript SK-
- Plasmid pMF6 (contain nos3' terminator)
- 25 Plasmid pHKH1 (contain maize adh1 intron)
- Plasmid MstsI(6-4) (contain maize stsI transit peptide, use as a template for PCT stsI transit peptide out)
- Plasmid MstsIII in pBluescript SK-
- Primers EXS29 (GTGGATCCATGGCGACGCCCTCGGCCGTGG) [SEQ ID NO:22]
- 30 EXS35 (CTGAATTCCATATGGGGCCCCCTCCCTGCTCAGCTC) [SEQ ID NO:23]
- both used for PCT stsI transit peptide
- Primers EXS31 (CTCTGAGCTCAAGCTTGCTACTTTCTTTTCCTTAATG) [SEQ ID NO:24]

EXS32 (GTCTCCGCGGTGGTGTCTTGCTTCCTAG) [SEQ ID NO:25]

both used for PCR maize 10KD zein promoter (Journal: Gene 71:359-370 [1988])
Maize A632 genomic DNA (used as a template for PCR maize 10KD zein promoter).

Step 1: Clone maize 10KD zein promoter in pBluescriptSK-(named as pEXS10zp).

- 5
1. PCR 1.1Kb maize 10KD zein promoter
primers: EXS31, EXS32
template: maize A632 genomic DNA
 2. Clone 1.1Kb maize, 10KD zein promoter PCR product into pBluescript SK-plasmid at SacI and SacII site (See FIG. 7).

10 Step 2: Delete NdeI site in pEXS10zp (named as pEXS10zp-NdeI).

NdeI is removed by fill in and blunt end ligation from maize 10KD zein promoter in pBluescriptSK.

Step 3: Clone maize adh1 intron in pBluescriptSK- (named as pEXSadhl).

15 Maize adh1 intron is released from plasmid pHKH1 at XbaI and BamHI sites. Maize adh1 intron (XbaI/BamHI fragment) is cloned into pBluescriptSK- at XbaI and BamHI sites (see FIG. 7).

Step 4: Clone maize 10KD zein promoter and maize adh1 intron into pBluescriptSK-(named as pEXS10zp-adh1).

20 Maize 10KD zein promoter is released from plasmid pEXS 10zp-NdeI at SacI and SacII sites. Maize 10KD zein promoter (SacI/SacII fragment) is cloned into plasmid pEXSadhl (contain maize adh1 intron) at SacI and SacII sites (see FIG. 7).

Step 5: Clone maize nos3' terminator into plasmid pEXSadh1 (named as pEXSadh1-nos3').

Maize nos3' terminator is released from plasmid pMF6 at EcoRI and HindIII sites.

Maize nos3' terminator (EcoRI/HindIII fragment) is cloned into plasmid pEXSadh1 at EcoRI and HindIII (see FIG. 7).

Step 6: Clone maize nos3' terminator into plasmid pEXS10zp-adh1 (named as pEXS10zp-adh1-nos3').

Maize nos3' terminator is released from plasmid pEXSadh1-nos3' at EcoRI and ApaI sites. Maize nos3' terminator (EcoRI/ApaI fragment) is cloned into plasmid pEXS10zp-adh1 at EcoRI and ApaI sites (see FIG. 7).

Step 7: Clone maize STSI transit peptide into plasmid pEXS10zp-adh1-nos3' (named as pEXS33).

1. PCR 150bp maize STSI transit peptide
primer: EXS29, EXS35

template: MSTSI(6-4) plasmid

2. Clone 150bp maize STSI transit peptide PCR product into plasmid pEXS10zp-adh1-nos3' at EcoRI and BamHI sites (see FIG. 7).

Step 8: Site-directed mutagenesis on maize STSI transit peptide in pEXS33 (named as pEXS33(m)).

There is a mutation (stop codon) on maize STSI transit peptide in plasmid pEXS33. Site-directed mutagenesis is carried out to change stop codon to non-stop codon. New plasmid (containing maize 10KD zein promoter, maize STSI transit peptide, maize adh1 intron, maize nos3' terminator) is named as pEXS33(m).

Step 9: NotI site in pEXS33(m) deleted (named as pEXS50).

NotI site is removed from pEXS33 by NotI fillin, blunt end ligation to form pEXS50 (see FIG. 8).

Step 10: Maize *adh1* intron deleted in pEXS33(m) (named as pEXS60).

5 Maize *adh1* intron is removed by NotI/BamHI digestion, filled in with Klenow fragment, blunt end ligation to form pEXS60 (see FIG. 9).

Step 11: Clone maize STSIII into pEXS50, pEXS60.

10 Maize STSIII is released from plasmid maize STSIII in pBluescript SK- at NdeI and EcoRI sites. Maize STSIII (NdeI-EcoRI fragment) is cloned into pEXS50, pEXS60 separately, named as pEXS51, pEXS61 (see FIGS. 8 and 9, respectively).

Step 12: Clone the gene in Table 8 into pEXS51 at NdeI/NotI site to form pEXS52. Other similar plasmids can be made by cloning other genes (STSI, II, WX, *glgA*, *glgB*, *glgC*, BEI, BEII, etc.) into pEXS51, pEXS61 at NdeI/NotI site.

15 Plasmid EXS52 was transformed into rice. The regenerated rice plants transformed with pEXS52 were marked and placed in a magenta box.

20 Two siblings of each line were chosen from the magenta box and transferred into 2.5 inch pots filled with soil mix (topsoil mixed with peat-vermiculite 50/50). The pots were placed in an aquarium (fish tank) with half an inch of water. The top was covered to maintain high humidity (some holes were made to help heat escape). A thermometer monitored the temperature. The fish tank was placed under fluorescent lights. No fertilizer was used on the plants in the first week. Light period was 6 a.m.-8 p.m., minimum 14 hours light. Temperature was minimum 68°F at night, 80°-90°F during the day. A heating mat was used under the fish tank to help root growth when necessary. The plants stayed in the

above condition for a week. (Note: the seedlings began to grow tall because of low light intensity.)

After the first week, the top of the aquarium was opened and rice transformants were transferred to growth chambers for three weeks with high humidity and high light intensity.

5 Alternatively, water mix in the greenhouse can be used to maintain high humidity. The plants grew for three weeks. Then the plants were transferred to 6-inch pots (minimum 5-inch pots) with soil mix (topsoil and peat-Vet, 50/50). The pots were in a tray filled with half an inch of water. 15-16-17 (N-K-P) was used to fertilize the plants (250 ppm) once a week or according to the plants' needs by their appearances. The plants remained in 14 hours
10 light (minimum) 6 a.m.-8 p.m. high light intensity, temperature 85°-90°/70°F day/night.

The plants formed rice grains and the rice grains were harvested. These harvested seeds can have the starch extracted and analyzed for the presence of the ligated amino acids C, V, A, E, L, S, R, E [SEQ ID NO:27] in the starch within the seed.

Example Seven:

15 SER Vector for Plants:

The plasmid shown in Figure 6 is adapted for use in monocots, i.e., maize. Plasmid pEXS52 (FIG. 6) has a promoter, a transit peptide (from maize), and a ligated gene fragment (TGC GTC GCG GAG CTG AGC AGG GAG) [SEQ ID NO:26] which encodes the amino acid sequence C V A E L S R E [SEQ ID NO:27].

20 This gene fragment naturally occurs close to the N-terminal end of the maize soluble starch synthase (MSTSI) gene. As is shown in TABLE 8, at about amino acid 292 the SER from the starch synthase begins. This vector is preferably transformed into a maize host. The transit peptide is adapted for maize so this is the preferred host. Clearly the transit peptide and the promoter, if necessary, can be altered to be appropriate for the host plant
25 desired. After transformation by "whiskers" technology (U.S. Patent Nos. 5,302,523 and 5,464,765), the transformed host cells are regenerated by methods known in the art, the

transformant is pollinated, and the resultant kernels can be collected and analyzed for the presence of the peptide in the starch and the starch granule.

The following preferred genes can be employed in maize to improve feeds: phytase gene, the somatotrophin gene, the following chained amino acids: AUG AUG AUG AUG
5 AUG AUG AUG AUG [SEQ ID NO:28]; and/or, AAG AAG-AAG AAG AAG AAG AAG
AAG AAG AAG AAG AAG {SEQ ID NO:29}; and/or AAA AAA AAA AAA AAA AAA
[SEQ ID NO:30]; or a combination of the codons encoding the lysine amino acid in a chain
or a combination of the codons encoding both lysine and the methionine codon or any
combination of two or three of these amino acids. The length of the chains should not be
10 unduly long but the length of the chain does not appear to be critical. Thus the amino acids
will be encapsulated within the starch granule or bound within the starch formed in the starch-
bearing portion of the plant host.

This plasmid may be transformed into other cereals such as rice, wheat, barley, oats,
sorghum, or millet with little to no modification of the plasmid. The promoter may be the
15 *waxy* gene promoter whose sequence has been published, or other zein promoters known to
the art.

Additionally these plasmids, without undue experimentation, may be transformed into
dicots such as potatoes, sweet potato, taro, yam, lotus cassava, peanuts, peas, soybean, beans,
or chickpeas. The promoter may be selected to target the starch-storage area of particular
20 dicots or tubers, for example the patatin promoter may be used for potato tubers.

Various methods of transforming monocots and dicots are known in the industry and
the method of transforming the genes is not critical to the present invention. The plasmid can
be introduced into *Agrobacterium tumefaciens* by the freeze-thaw method of An et al. (1988)
Binary Vectors, in Plant Molecular Biology Manual A3, S.B. Gelvin and R.A. Schilperoot,
25 eds. (Dordrecht, The Netherlands: Kluwer Academic Publishers), pp. 1-19. Preparation of
Agrobacterium inoculum carrying the construct and inoculation of plant material, regeneration
of shoots, and rooting of shoots are described in Edwards et al., "Biochemical and molecular
characterization of a novel starch synthase from potatoes," Plant J. 8, 283-294 (1995).

A number of encapsulating regions are present in a number of different genes. Although it is preferred that the protein be encapsulated within the starch granule (granule encapsulation), encapsulation within non-granule starch is also encompassed within the scope of the present invention in the term "encapsulation." The following types of genes are useful for this purpose.

Use of Starch-Encapsulating Regions of Glycogen Synthase:

E. coli glycogen synthase is not a large protein: the structural gene is 1431 base pairs in length, specifying a protein of 477 amino acids with an estimated molecular weight of 49,000. It is known that problems of codon usage can occur with bacterial genes inserted into plant genomes but this is generally not so great with *E. coli* genes as with those from other bacteria such as those from *Bacillus*. Glycogen synthase from *E. coli* has a codon usage profile much in common with maize genes but it is preferred to alter, by known procedures, the sequence at the translation start point to be more compatible with a plant consensus sequence:

glgA G A T A A T G C A G [SEQ ID NO:31]
cons A A C A A T G G C T [SEQ ID NO:32]

Use of Starch-Encapsulating Regions of Soluble Starch Synthase:

cDNA clones of plant-soluble starch synthases are described in the background section above and can be used in the present invention. The genes for any such SSTS protein may be used in constructs according to this invention.

Use of Starch-Encapsulating Regions of Branching Enzyme:

cDNA clones of plant, bacterial and animal branching enzymes are described in the background section above and can be used in the present invention. Branching enzyme [1,4Dglucan: 1,4Dglucan 6D(1,4Dglucano) transferase (E.C. 2.4.1.18)] converts amylose to amylopectin, (a segment of a 1,4Dglucan chain is transferred to a primary hydroxyl group in a similar glucan chain) sometimes called Q-enzyme.

The sequence of maize branching enzyme I was investigated by Baba et al. (1991) BBRC, 181:87-94. Starch branching enzyme II from maize endosperm was investigated by

Fisher et al. (1993) Plant Physiol, 102:1045-1046. The BE gene construct may require the presence of an amyloplast transit peptide to ensure its correct localization in the amyloplast. The genes for any such branching enzyme of GBSTS protein may be used in constructs according to this invention.

5 **Use of Starch-Binding Domains of Granule-Bound Starch Synthase:**

The use of cDNA clones of plant granule-bound starch synthases are described in Shure et al. (1983) Cell 35:225-233, and Visser et al. (1989) Plant Sci. 64(2):185-192. Visser et al. have also described the inhibition of the expression of the gene for granule-bound starch synthase in potato by antisense constructs (1991) Mol. Gen. Genetic 225(2):289-296; 10 (1994) The Plant Cell 6:43-52.) Shimada et al. show antisense in rice (1993) Theor. Appl. Genet. 86:665-672. Van der Leij et al. show restoration of amylose synthesis in low-amylose potato following transformation with the wild-type waxy potato gene (1991) Theor. Appl. Genet. 82:289-295.

The amino acid sequences and nucleotide sequences of granule starch synthases from, 15 for example, maize, rice, wheat, potato, cassava, peas or barley are well known. The genes for any such GBSTS protein may be used in constructs according to this invention.

Construction of Plant Transformation Vectors:

Plant transformation vectors for use in the method of the invention may be constructed using standard techniques

20 .

Use of Transit Peptide Sequences:

Some gene constructs require the presence of an amyloplast transit peptide to ensure correct localization in the amyloplast. It is believed that chloroplast transit peptides have similar sequences (Heijne et al. describe a database of chloroplast transit peptides in (1991) Plant Mol. Biol. Reporter, 9(2):104-126). Other transit peptides useful in this invention are 25 those of ADPG pyrophosphorylase (1991) Plant Mol. Biol. Reporter, 9:104-126), small subunit RUBISCO, acetolactate synthase, glyceraldehyde3Pdehydrogenase and nitrite reductase.

The consensus sequence of the transit peptide of small subunit RUBISCO from many genotypes has the sequence:

MASSMLSSAAVATRTNPAQASM VAPFTGLKSAAFPVSRKQNLDI TSIASNGGRVQC
[SEQ ID NO:33]

5 The corn small subunit RUBISCO has the sequence:

MAPTVMMASSATATRTNPAQAS AVAPFQGLKSTASLPVARRSSR SLGNVASNGGRIRC
[SEQ ID NO:34]

The transit peptide of leaf glyceraldehyde3Pdehydrogenase from corn has the sequence:

10 MAQILAPSTQWQMRITKTSPCA TPITSKMWSSLVMKQTKKVAHS
AKFRVMAVNSENGT [SEQ ID NO:35]

The transit peptide sequence of corn endosperm-bound starch synthase has the sequence:

15 MAALATSQLVATRAGHGVDPDASTFRRGAAQGLRGARASAAADTLSMRTSARAAPRHQ
QQARRGGRFPFPSLVVC [SEQ ID NO:36]

The transit peptide sequence of corn endosperm soluble starch synthase has the sequence:

MATPSAVGAACLLLARXAWPAAVGDRARPRRLQRVLRRR [SEQ ID NO:37]

Engineering New Amino Acids or Peptides into Starch-Encapsulating Proteins:

20 The starch-binding proteins used in this invention may be modified by methods known to those skilled in the art to incorporate new amino acid combinations. For example,

sequences of starch-binding proteins may be modified to express higher-than-normal levels of lysine, methionine or tryptophan. Such levels can be usefully elevated above natural levels and such proteins provide nutritional enhancement in crops such as cereals.

5 In addition to altering amino acid balance, it is possible to engineer the starch-binding proteins so that valuable peptides can be incorporated into the starch-binding protein. Attaching the payload polypeptide to the starch-binding protein at the N-terminal end of the protein provides a known means of adding peptide fragments and still maintaining starch-binding capacity. Further improvements can be made by incorporating specific protease cleavage sites into the site of attachment of the payload polypeptide to the starch-encapsulating
10 region. It is well known to those skilled in the art that proteases have preferred specificities for different amino-acid linkages. Such specificities can be used to provide a vehicle for delivery of valuable peptides to different regions of the digestive tract of animals and man.

In yet another embodiment of this invention, the payload polypeptide can be released following purification and processing of the starch granules. Using amylolysis and/or
15 gelatinization procedures it is known that the proteins bound to the starch granule can be released or become available for proteolysis. Thus recovery of commercial quantities of proteins and peptides from the starch granule matrix becomes possible.

In yet another embodiment of the invention it is possible to process the starch granules in a variety of different ways in order to provide a means of altering the digestibility of the
20 starch. Using this methodology it is possible to change the bioavailability of the proteins, peptides or amino acids entrapped within the starch granules.

Although the foregoing invention has been described in detail by way of illustration and example for purposes of clarity and understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and
25 modifications may be made thereto without departing from the spirit or scope of the appended claims.